Machine Decision Is Not Final

China and the History and Future of Artificial Intelligence

机器决策并非终局

中国与人工智能的历史及未来

Edited by
Benjamin Bratton
Anna Greenspan
Amy Ireland
Bogna Konior

Blaise Agüera y Arcas
Bo An
Benjamin Bratton
Chen Qiufan
Gabriele de Seta
Shuang L. Frost
Vincent Garton
Steve Goodman
Yvette Granata
Anna Greenspan
Amy Ireland
Bogna Konior
Vincent Lê
Lawrence Lek
Lukáš Likavčan
Suzanne Livingston
Iris Long
Bingchun Meng
Reza Negarestani
Wang Hongzhe
Regina Kanyu Wang
Xin Wang
Xia Jia
You Mi

Machine Decision Is Not Final

China and the History and Future of Artificial Intelligence

机器决策并非终局

中国与人工智能的历史及未来

Edited by
Benjamin Bratton
Anna Greenspan
Amy Ireland
Bogna Konior

URBANOMIC

Published in 2025 by Urbanomic Media Ltd,
The Old Lemonade Factory,
Windsor Quarry,
Falmouth TR11 3EX
United Kingdom

Publication of this volume was supported by Research Center for Artificial Intelligence and Culture, NYU Shanghai.

© Urbanomic Media Ltd.
All rights reserved.
No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any other information storage or retrieval system, without prior permission in writing from the publisher.

British Library cataloguing-in-publication data
A full catalogue record of this book is available from the British Library
ISBN (Print Edition) 978-1-913029-99-9
Distributed by the MIT Press, Cambridge, Massachusetts and London, England

www.urbanomic.com

d_r0

Anna Greenspan and Bogna Konior

Introduction: Fleeting Forces and Clever Machinations

> *I've heard my teacher say, 'Where there are clever machines, there will necessarily be clever machinations, and where there are clever machinations, there will necessarily be mechanical hearts and minds.' Once the mechanical heart is lodged in your breast, purity and simplicity are no longer complete in you. When purity and simplicity are no longer complete, the imponderable spirit and life in you become unsettled. When the spirit and life in you are unsettled the Way cannot be carried by them. It's not that I don't know about this thing, it's that I would be ashamed to use it.*
>
> ZHUANGZI[1]

> *Unless the whole universe is in the process of fusion and intermingling like fleeting forces moving in all directions, it may not be called The Great Harmony.*
>
> ZHANG ZAI[2]

From at least as early as the Warring States Period (475–221 BCE) Chinese philosophy has been concerned with the question of machines. The great intellectual flourishing of 'a hundred schools of thought' during this period coincided with widespread social and technological change. Descriptions of imaginary self-driving carriages, clever mechanical birds sent out to spy on enemies, and other fantastical contraptions provoked worries about destruction beyond human control in an epoch already marred by war.[3] Early Confucian and Daoist thinkers vigorously debated the nature of artifice. Is following and imitating nature the only right path? What is the role of creative innovation? Are technological developments inherently disruptive to the maintenance of social relations and the work of retaining a harmonious relationship between human beings and the cosmic order?[4] Or might the ever-changing *Dao* (道), or Way, be entirely consistent with the evolution of machines? Attitudes to technology dating from this pre-modern period range widely and, despite their remoteness in time, mirror recurrent concerns found, for example, at the dawn of the Industrial Revolution in Europe. Today, they resonate acutely with the concerns of the present, as artificial intelligence reaches a critical juncture, moving out of the laboratory and into the fabric of everyday life. Inside and outside of China,

two major questions return cyclically, as if proper to any period of technological upheaval. First, what is a machine? Is it part of cosmic logic and the way of nature, or does it stand outside of them? Does it belong among the 'ten thousand things' of the *Zhuangzi* that all share in the *Dao*, or is it an artificial abomination, a fundamental discontinuity or rupture in the cosmic continuity linking heavens, earth, and humans—*tiandiren* (天地人)?[5] Second, is there a capacity for decision or an agency proper to a machine? Is a machine simply a tool for humans to use, or does it have its own way of acting that is irreducible to human designs? Can it be tamed, or is it inherently threatening to human civilisation? Does technology necessarily reflect human goals, or does it channel a purpose beyond human dominion? Is there a machinic will that does not simply follow the will of its user? Is technology under our control or, alternatively, are we controlled by our technologies? In this book we also want to ask whether China's philosophical and historical engagement with these questions sheds light on the destiny of machines in human history as such—outside the confines of any particular culture or territory. We do not wish to distil a single 'Chinese' philosophy of artificial intelligence or determine which form of thought most accurately captures China's present, past, or future. We do not aspire to 'represent' or 'explain' China. Rather, our contention is both simple and obvious: one cannot properly tune into the currents of machine intelligence flowing across the planet if one remains confined within a single geographical perspective.

That certain concerns about technology are, for lack of a better word, universal, should be evident to any historian. Early Daoists cautioned against technologies that sully the spirit and ignite chaos. The *Zhuangzi* warns against the use of 'clever machines' and suggests that they are better left alone by those walking the righteous path.[6] This is not unlike the apprehension that many European intellectuals harboured toward the technologies of their respective times, whether that be the early widespread resistance in Western Europe to Hindu-Arab numerals, the reluctance of the Catholic Church to embrace perspectival painting and scientific method in the early Renaissance, or the anxiety that swept over the European continent during the introduction of photography, the film camera, the microscope, and the steam engine.[7] 'To try to capture fleeting images is not just an impossible undertaking, the very wish to do such a thing is blasphemous', reads a nineteenth-century German critique of early daguerreotype

technology. ‘Man is made in the image of God, and God’s image cannot be captured by any machine.’[8] Yet admonition and derision are far from being the only reactions. Humanity’s enthusiastic embrace of technology, which is so patent in practice, can also be found in the realm of ideas.

In China, Mohists took an artisan position, forcefully arguing against the Confucian ideal to transmit rather than create.[9] The Neo-Confucian Shen Kuo (1031–1095), whose *Brush Talks from Dream Brook* documents scientific innovations in physics, astrology, mathematics, and medicine, claimed that there was no contradiction between Confucius’s teachings and technology.[10] Despite the apparent Daoist rejection of machines, there is arguably a Daoist attitude that views the emergence of technology as consistent with the Way. The ideal of *wuwei* (无为, effortless action) is commonly illustrated by the famous story of Cook Ding, who learned to follow the hollows and openings of an animal’s body and effortlessly wielded his knife to carve an ox.[11] Such ‘effortless action’ is arguably aligned with bottom-up trends in machine learning and recent discoveries in neuroscience regarding the spontaneous intelligence of nonconscious cognition.[12] Perhaps the most significant contribution to any ‘pro-machine’ tendency in Chinese thought, however, comes from the profound influence of Buddhism. Buddhist culture and ideas, which spread across Asia through the trade routes and communication networks of an early urban cosmopolitanism, have long had a deep affinity with technoculture.[13] Buddhists were amongst the first adopters of print technology, and have since then actively incorporated a wide range of ‘technologies of salvation’ including photographs, WeChat dharmas, smartphone rituals, and robot monks.[14] Buddhist modernity made a deep impact on thinkers such as Tan Sitong who saw in the emerging world of electromagnetic communication an expression of New Confucian philosophy.[15]

These traditions are occasionally called upon to justify why China, according to some studies, scores high on the social acceptance of robotics and low on AI fear.[16] To explain why ‘China does not fear AI’, for example, Bing Song draws upon schools of thought whose early practitioners often rebuked machines; she writes that ‘notions of living with other intelligent beings, most of which are more powerful and long-lasting than humans, abound in Daoist and Buddhist thinking’.[17] Indeed, the Mahayana conception that even non-sentient beings have Buddha Nature seems to resonate with the possibility of ‘artificial’ intelligences. Gai Fei points out

that, contrary to a perceived organic prejudice amongst Daoists, practitioners of inner alchemy actively rejected a plant-based diet in favour of ingesting more durable materials such as crystals and metals.[18] In their quest for immortality, they sought to create a metallic body, enacting a kind of proto-cyborgian becoming. Confucianism has also been called upon to accommodate artificial intelligence: contemporary philosophers such as Li Chenyang and Yao Zhongqiu argue that, rather than threatening harmonious human social order, AIs may belong within the Confucian ethic of 'graded love' and be counted among the creatures (*wu*, 物) with which we partake in friendly and mutually beneficial relationships.[19] Likewise, Stephan Angle proposes that the capacity to program software, which can encourage certain virtues, suggests that the ubiquity of AI may be of use in constructing a society with Confucian values.[20]

Yet attitudes and narratives around artificial intelligence both inside and outside of China are ambivalent and contradictory. Arguments suggesting an inherent sympathy between AI and traditional Chinese philosophy can easily serve national business interests and bolster geopolitical aspirations. A few years ago, China 'plunged into an artificial intelligence fever', as enthusiastically described by the well-known executive and investor, Kai-Fu Lee.[21] In 2016, more than 280 million Chinese viewers tuned in to watch Google's AlphaGo defeat Lee Sedol in the ancient game of Go. In 2018, the bookshelves behind President Xi Jinping in his New Year speech displayed two AI-related publications, Pedro Domingos's *The Master Algorithm* and Brett King's *Augmented: Life in the Smart Lane*, and Xinhua News Agency unveiled its new AI anchors, Qiu Hao and Xin Xiaomeng.[22] There are hopes amongst both government officials and corporate leaders that AI and robotics can finally resolve the flaws of a centralised planned economy.[23] For some, the future of the party lies, ultimately, in the realisation of a 'fully automated luxury communism'[24] with Chinese characteristics. Yet, despite a government declaration that China would become a leading AI power by 2030 and the optimistic prognosis of entrepreneurs such as Lee, and even accounting for disruptive breakthroughs from China, it is still in America that the most advanced, cutting-edge research takes place, and it is still American technoculture that reigns supreme in the popular imagination.

There is a common accusation that China—even at its most inventive—ultimately only copies innovations from elsewhere. This corresponds to a long-held argument that technoscientific advances in China are embedded

in the ideology of ‘scientism’, which took root during the May 4th period and advocated for the rejection of Chinese traditions in favour of a wholesale adoption of what were deemed to be more progressive ‘Western’ ideas. As Shiping Hua writes, the doctrine of scientism implies that science is unified; has no limits; has been extremely successful at prediction, explanation and control; confers objectivity, and is beneficial as a whole.[25] Although the relationship between science as a way of knowing the world and technology in its multiple functions is not so obvious, ‘science and technology’ can serve as relatively uncontroversial ideals to rally behind, and are easily tied into the narrative of economic and civilisational progress. These competing conceptions of the relationship between culture, philosophy, and technology are actively debated as ‘Chinese intellectuals [struggle] to fill an intellectual vacuum created by the failure of the Cultural Revolution [and] the subsequent gradual loss of influence of Marxism as a whole, and to re-vision a polity for China’.[26]

It is true that one of the most common responses in China to technologies such as doorways and vending machines that operate by facial recognition is to remark upon their convenience,[27] an overly simplistic techno-optimistic narrative obscures the fact that Chinese citizens are as conscious and weary of the ethical and political problems of AI as their counterparts elsewhere in the world. Algorithmic control of the cityscape is a widespread concern, while ‘algorithmic bias, discriminative uses, responsibility and legal issues’ as well as ‘concerns over job displacements caused by AI’, ‘rising inequality’, and ‘the loss of autonomy’ are repeatedly discussed on WeChat and Zhihu.[28] Predictably, as Jinghan Zeng notes, objections to AI may be attributed to the lack of legal recourse to methods of resisting surveillance. There is also substantial concern that, owing to its demographics and stage of economic development, AI’s transformation of the workforce may prove particularly challenging in China.[29]

Finally, a positive narrative of AI as an extension of ancient Chinese wisdom or a carrier of civilisational progress stands in contrast to more recent government policy. Although Chinese military and industrial sectors are still prioritising AI,[30] economic and data privacy concerns, as well as the Chinese state’s larger desire to construct an oppositional ‘non-Western’ identity, means that American software, including generative AI, is as of now prohibited and requires a VPN to access. And while Chinese AI

models are proliferating, these are subject to strict censorship and risk being blocked by uncertain regulation.[31]

Mapping connections between Chinese thought and the question of the machine consequently reveals an adjacent area of investigation: What is the relationship between technologies and the territories whence they originate and in which they are used? Are particular technologies tied to specific cultures or cosmologies? Is technological development historically and culturally contingent? Can certain machines only emerge in specific times and places, under specific cultural conditions? To state it most broadly, what is the relationship between 'culture' and 'technology'? On one hand, the answers to these questions would seem straightforward. Isn't technology simply a subset of culture? And yet, if this were the case, shouldn't there be as much technological diversity as there is cultural diversity? Does the apparent lack of diversity suggest that the trajectory of technological development is universal or, on the contrary, does modern technocapitalism bear the unique characteristics of 'Western culture' imposed the world over such that, as thinkers such as Weber and Marx predicted, 'modernisation' is always essentially 'Westernisation'? Alternatively, is technology itself created from older layers of cultural exchange, impossible to pin to one territory or another?

This latter line of thought, that technologies emerge out of networks of cultural exchange, is supported by the influence of the cross-cultural transmission of the *Yijing*, or *The Book of Changes*, on the history of computation. This ancient compendium, which can be read as a cosmological treatise, an early scientific theory pertaining to natural laws, a book of advice and wise sayings, and a manual for divination, was first introduced to Europe by the Jesuits in the early modern period. It was of particular interest to polymath philosopher Gottfried Leibniz who held that the discovery of binary arithmetic could be found in the numeracy of the *Yijing*'s hexagrams and used it as inspiration for his early calculating machines.[32] In addition to its transmutations into binary code in industrialising Europe, the *Yijing* sits at the cusp between prophecy and prediction and is thus foundational to the history of digital computation. Notwithstanding these intersecting histories of technological cultures, the problem of the exact relationship between technology and territory continues to animate philosophy.

Since the Industrial Revolution, there has existed a definition of technology as a realm supposedly purged of cosmology, and related instead to a secular global modernity. In *The Question Concerning Technology in China: An Essay in Cosmotechnics*, Yuk Hui argues that we suffer from a lack of technodiversity which can only be countered by isolating the exact ontology of 'Western technology' as well as producing a completely distinct one based in Chinese philosophy, or other 'non-Western' philosophies.[33] The idea of technology prevalent in the West from ancient Greek thought until now, he argues, was identified by Heidegger: the use of tools to mould the world into forms suitable to human goals, needs, and designs. The dominance of this idea alongside occidental 'Promethean' modernity and colonialism is such that it has become nearly impossible to even imagine the 'question of technology' differently. This is not a simple cultural difference, but rather a metaphysical one, implying a very specific relationship between humans and the cosmos, where the world is a 'standing reserve' for human use, and *techne* is understood simply as a set of devices through which humans shape the world in accordance with their needs and desires.[34] This fundamental homogenisation of technology, Hui argues, can only be challenged by a completely distinct cosmological and technical practice, or 'cosmotechnics', that emerges outside the philosophical culture of the West. 'In the West', he writes, 'there is no longer the concept of cosmotechnics because there is no longer cosmology but only astrophysics'. The question of technology in China, Hui goes on to suggest, which is based on the concepts *qi* (器, tools) and *Dao* (道, [the] Way), may still give rise to a truly alternative technoculture that is expressive of the Way and capable of uniting heavens, earth, and humans.[35]

Hui is not alone in his desire to conceive of alternative technocultures, yet other than the example of Chinese medicine,[36] which is rooted in a particular understanding of the body and its expressive capacities,[37] it is difficult to imagine concrete examples of what these other technologies might be. You Mi offers a provocative example in her analysis of Liu Cixin's science fiction novel *The Wandering Earth*, in which the Earth in its entirety is turned into a spaceship. In contrast to the common notion of a spaceship as a tool for leaving the Earth in order to colonise other planets, she writes that 'every Chinese audience can relate this to the idea of 安土重迁 (*antu chong qian*), meaning [...] attachment to one's homeland to the point of unwillingness to leave it behind, a deeply internalized idea thanks to the

traditional Confucian and agrarian social order'.[38] Similarly, in their projects *How to Eat and Code in Chinese: Sinofuturist Recipes* (2019) and *How to Feed an AI* (2020), Xiaowei R. Wang speculates about types of automation and intelligence that do not privilege the brain but, drawing on Chinese medicine and culinary culture, the nourishment of the senses.[39] Departing from the argument that Chinese medicine presents a distributed and decentralised sensory model in which eleven organs are holistically involved in circulating intelligence throughout the body, Wang imagines a scenario where a team of Chinese scientists at Alibaba AI 'started from Chinese theories of the body, Chinese medicine, and Buddhist thought […] to create a pretrained neural network with electrical inputs and outputs from a combination of artificial and human organs', a hybrid machine that does not need to ingest data.[40] Still another example can be found in Lawrence Lek's artworks: *Geomancer* (2017) envisions an AI that reaches enlightenment through its exposure to Buddhist chants and Daoist philosophy, while *Black Cloud* (2021) and *NOX* (2023–2024) portray a smart city in the near future, where it is not certain whether machine intelligence is a question of consciousness or legal liability, uniting cosmological and bureaucratic trajectories of thought.[41]

These varied speculations at times recollect the history of AI in China, which, whilst deeply connected with global developments, is also riddled with tangents that follow their own singular path. Thomas Mullaney argues that adapting the Chinese writing system, with its thousands of unique characters, to modern media such as the telegraph, typewriters, and digital computers, involved novel innovations in human-machine interaction that prefigured predictive text. This process of conversion, especially between Chinese characters and the alphanumerics of the Qwerty keyboard, requires a fundamental arbitrariness and hypermediation which, Mullaney contends, results in a new form of writing he calls 'hypography' which, he believes, is profoundly attuned to the real, underlying 'plasticity, of all humanmachine interaction'.[42] Meanwhile, as Xiao Liu documents, China joined the 'information craze'—synonymous with breakthroughs in both Japan and the US in so-called 'expert systems'—in the 1980s, when the country was steeped in 'qigong fever'—*qigong re* (气功热).[43] The prospect of connecting computation with *qi*-based cosmology mobilised some of China's greatest minds, including Qian Xuesen, whose singular vision for the future of computation was based on a synthesis of somatic and technical

sciences. Another intellectual who challenged the correlation between computation and universalism was the Chinese-American logician Hao Wang. In the 1950s, he drew upon Eastern Marxism and Western analytic philosophy to critique the then conventional view of cognitive processes as mechanical computation, arguing instead that human reasoning is personal and influenced by historical and material conditions rather than universal rules.[44] More recently, Midjourney founder David Holz, when describing his experience with DeepSeek, noted that, compared to Western models, it excelled in Chinese philosophy and literature while retaining strong English language capabilities. He speculated that this is thanks to China's well-preserved literary history in contrast to that of the West, which has lost many of its ancient texts. R1, he added, has a different philosophical foundation, potentially opening unique paths for AI self-actualisation.[45] It is still unclear, however, whether these alternative histories and poetic interventions translate into a real, concrete diversity in AI. Rather, at least for the time being, gestures toward Chinese cosmology often seem to function as little more than rhetorical dressing on accounts of all-too-familiar technologies. For example, Alibaba's Aspara databases, named after the ethereal Buddhist nymph, store a wealth of information necessary for the company's proposed project ET Agricultural Brain in which millions of pigs are optimised and produced at a low price, while farmers become managers or administrators.[46]

Cases such as these may suggest that it is not Confucianism, Buddhism, or Daoism, but rather China's long tradition of legalism and respect for bureaucracy that underlies its engagement with AI. It might even be argued that 'digital technology in contemporary China emerged from the need to rebuild a new form of bureaucracy'.[47] If seen in isolation, China's enthusiasm for technocratic bureaucracy supports the popular image of Chinese technology as focused on control and synchronisation, whether that takes the form of a 'techno-surveillance state' or the legacy of communism's insistence on the 'mass'.[48] In the worst cases, as Benjamin Bratton notes in this volume's afterword, when seen from outside China, these narratives recast the country's inhabitants as robots in a techno-dystopian reimagining of 'the yellow horde', where Sinophobia and technophobia converge into a general fear of the future pictured as 'an emotionless mode of total computational control, with distinctly Asian characteristics' soon to replace 'humanity'.[49] From efficient infrastructure

development and smoothly-run biometric cities to choreographed dancing robots and its response to COVID-19, China has been both admired and scorned for its aptitude for technological coordination in the service of an 'authoritarian resilience'.[50] In the period immediately after the pandemic, the idea that 'acceleration is both an addiction and necessity for the Shenzhen model' was severely questioned.[51] The 'Sinofuturist' proclamation that 'neo-China arrives from the future'[52] as a designation of a different world order to come coexisted with a post-pandemic malaise that has been dubbed 'Sino-no-futurism' and 'Sinopessimism'.[53] The mood was also sour in online youth subcultures, as indicated by the take-off in early 2021 of *tangping* (躺平)—the doctrine of 'lying flat'[54]—or a video posted on Weixin in May 2022 in which a young couple face off against a trio of *dabai* (大白), or 'big whites' (workers clad in white hazmat suits enforcing the government's Zero Covid policy). In response to the couple's refusal to be taken to a quarantine centre, one of the *dabai* threatens them with punishment for being a menace to public security, adding ominously that it will affect 'the next three generations' of their family. The young couple's nihilistic response, uttered calmly before politely closing the door, saw the video go viral: 'Thanks, but we are the last generation.'[55] Yet, despite the growing power of the state, many of the essays in our book attempt to think AI in terms that are not just limited to power and control. We recognise that, rather than unfolding neatly from the top down according to an overarching plan, AI in China is driven by fragmented policies and a complicated dynamic that is based in part upon the substantial and well-established tensions between local and central governments, as well as the entrepreneurial, scientific, and creative energy of companies. There is constant negotiation between a shifting and diverse set of actors, both private and public, with competing interests, all of whom are vying for a limited set of resources. The state is working to balance a contradictory position of simultaneously relying on and reining in the country's tech giants, whilst AI has the potential to both bolster and weaken the legitimacy of the CCP.[56]

Instead of repeating common tropes and anxieties, this book focuses on the gaps and glitches, where countercultural practices in cyberspace, forgotten histories, and neglected philosophical lineages flourish. In China as elsewhere, the cybernetic fantasy of a flawless mechanistic society ruled by a hyperefficient synthetic autocrat is only ever imperfectly realised. This

tension was expressed in the summer of 2022, a few months before the frenzied response to OpenAI's release of ChatGPT, when the Shanghai lockdown had just been lifted but the country was still gripped by a Zero Covid policy that had morphed into a futile maze of convoluted and unpredictable procedures. On Chinese social media, news of the ever-stricter, constantly changing rules alternated with reports on events promoting China as a future technological utopia. The World Artificial Intelligence Conference at Shanghai's Zhangjiang Science Hall featured robot baristas and 'chairs printed by the most advanced metaverse printing machine in order to make NFTs accessible to everyone'.[57]

Image posted by parody account @neighbours_wifi on X in 2022.

A machine translation of a notice displayed at an Entry-Exit Bureau in China. This image circulated widely on Chinese social media during the Winter/Spring 2022 Covid lockdown.

The tension between the narratives of prophesied technological progress and the lived reality of digitised politics during lockdown was captured brilliantly in two signs photographed during this period. The first inspired the title for this volume: posted on a Twitter parody account in 2022, a machine-translated notice on a claw machine proclaims implacably that ‘Machine’s Decision is Final’. In the second, a notice announces apologetically that the country remains closed to foreigners: 外国人来华工作许可系统故障，请耐心等待. A literal translation would be ‘The work permit system for foreigners in China is malfunctioning, please wait patiently’. But it is the machine mistranslation, printed boldly in red by the unassuming administration, that went viral. Undermining any assumptions of hyperefficiency and automated finality, it requests politely that visitors ‘please wait patiently for the failure of the system’.

Although ‘artificial intelligence’ can refer both to existing technologies encompassing machine learning, image recognition software, neural networks, and generative adversarial networks, as well as the mythical goal

of an engineered nonhuman mind, the early 2020s have been defined by dramatic advances in Large Language Models. The way was initially paved by American companies including OpenAI, Alphabet, Microsoft, and Meta. The decades-old imaginary of intelligent machines skilled at conversation and creative tasks, as embodied in Alan Turing's 1950 paper 'Computing Machinery and Intelligence', seemed to have been borne out by American chatbots.[58] Initially, synthetic language and content creation apps were not the centre of attention in China, and only a few specific examples briefly captured the nation's imagination.[59] Additionally, the restrictions placed on giant firms such as Baidu, Tencent, and Alibaba have contributed to the perception that AI development in China has fallen behind. Yet the 2025 release of DeepSeek suggests that technological development is characterised by its own inherent wave-like rhythm, irrespective of the politics of the state.[60] That a relatively unknown player was able to create such an advanced model despite chip blockades reignited discussions about Chinese innovation in the US-dominated AI landscape.[61] It is worth noting that Chinese researchers constitute the majority of the world's top AI talent, working both domestically and internationally. In 2022, for instance, Chinese AI researchers were a larger presence in US institutions than American researchers.[62] According to the World Intellectual Property Organization (WIPO), China has published more patents in generative AI annually than all other countries combined since 2017.[63] China is also well-represented in most high-impact research publications in AI.

From 5G and TikTok to electric vehicles, *Black Myth Wukong*, and UniTree robots, Western responses to Chinese technological innovations oscillate between shock and sanctions. Rather than these creative achievements being recognised as part of a consistent pattern, they are often dismissed as isolated exceptions. Yet China's strident policies of cyber-sovereignty—its strict bans on popular global platforms including Google, YouTube, and X—has spawned a vibrant and singular cyberculture that has thrived in relative isolation, building a self-sufficient yet globally competitive ecosystem. The superiority of its algorithms in video streaming culture, for example, is evidenced by the power of TikTok, a platform seemingly dedicated to frivolous entertainment, which found itself at the centre of a Sino-American geopolitical struggle in early 2025. The fact that ByteDance (TikTok's parent company) is part of a larger, similarly robust mediasphere became clear to the 'Tiktok refugees' who flocked to

Xiaohongshu in protest against the threat of their favourite app being banned by the US government. The West's fears, scepticism, and punitive approach, then, has paradoxically helped to enable the development of a distinct and competitive Chinese media environment. The introduction of DeepSeek's R1 model caused such political and economic turmoil because it implied that Chinese AI might follow a similar trajectory.

Although there is some truth to the accusation that Chinese models are 'just copies' of more advanced innovations happening in the high-cost research labs of American companies, it is precisely this ability to rapidly reproduce high-tech products at a fraction of the cost that is at the heart of Chinese technoindustrial competitiveness. This was made clear by the *shanzhai* (copycat) mode of cell phone production centred in the coastal city of Shenzhen, which, during the first decades of the millennium, enabled the widespread proliferation of cheap mobile phones. DeepSeek demonstrated that this mode of production can also be mobilised in the realm of AI. By refining its algorithms to maximise computational efficiency, DeepSeek—operating under a chip ban—leveraged older, less powerful hardware through software innovations, reducing energy consumption and challenging the narrative that AI scaling must rely on exponential resources. While this benefits the development of AI as a whole, it also serves strategic national aims: it may reduce American AI companies' economic advantage and weaken the role of high-end chip manufacturing, once considered indispensable for further technological advances. Moreover, this achievement strengthens the potential of Hangzhou—home to DeepSeek and the tech giant Alibaba—to become the next Shenzhen: an epicentre of computer culture in the coming decades.[64]

More significantly, DeepSeek has renewed the promise of free, open-source technologies; once released into the world, these cannot be contained. This points toward a future of AI multiplicity where models may exist beyond the control of any single state or corporation. While an understanding of both Chinese and American technological cultures is essential for grasping AI, framing this solely as national competition misses a crucial point about how this technology unveils itself both within and outside of human histories. R1's most significant technical and philosophical implication is that its innovations rely on shifting humans further away from the development process. As this process accelerates beyond anthropocentric oversight, we may ultimately lose our grasp of what

these systems are optimising for. Like many of its counterparts abroad, DeepSeek is less interested in commercial applications than in the broader goal of building artificial general intelligence, and the company's tagline, *mai xiang wei zhi* (迈向未知)—'Into the Unknown'—is a fitting slogan for these aspirations.[65]

Artificial intelligence is not simply a medium in the sense of a tool for communication or an infrastructure for the facilitation of social and political management.[66] The history of technology seems to us to unfold through both the sociopolitics of human decisions and an irreducible outside logic which might only be perceptible through machinic or nonhuman channels. Rather than simply asking about culture and territory, we also wonder whether technology—just like mathematics or language—might sit uneasily between invention and discovery: What do we create and what is transmitted from outside? Rather than arguing in favour of either technodiversity or technouniversality, we see technology as encompassing an inhuman vector following a trajectory that traverses human activity without being reducible to it. At the limit we are interested in questions of time and agency. Are we cultivating AI or is it cultivating us? Who is training whom? What strange temporalities might be involved? Is it the human past that is determining the future of AI or might it be the other way around?

By proposing a logic which operates in an unclear relation to human decisions, we take a different path to scholarship that overemphasises technological development's 'non-Western' spirit or can only qualify its propositions by arguing differences from, or similarities to, an overarching 'West'. Such arguments often amount to running in circles: if an attempt is made to capture something local, it is 'essentialist and fetishistic'; if one points toward what is common, universal even, one is accused of perpetuating 'Western modes of thinking'. We might say that such logic is self-defeating, but perhaps it is so by design, because it preserves the current hierarchies that determine which authors get read and recognised as thinkers, and which are seen as merely an artefact of their 'national culture'. So much ink has been spilled on this subject that Xin Wang writes that 'the Other has gained a voice, but only to continuously explain, qualify and make sense of itself; [this] takes up much of the space necessary for […] operating outside the Euro-American epistemological comfort zone'.[67] Indeed, rarely, if ever, are books published on *British Philosophy of*

Computing in the 21st Century, or *Recent American Thought About AI*. Perpetually in search of its own Other, the centre is still invisible to itself and unmarked, while what it sees as the margins are denoted geographically as exotic regions of 'representative' thought still to be explored, often tasked with a redemptive mission of wresting AI away from what is perceived as a sinful entanglement with some kind of 'villain', be it 'the West', 'capitalism', or 'colonialism'. It is our contention that anyone interested in emerging technologies and artificial intelligence should be invested in what is happening beyond this landscape, not because of some preexisting interest in specific regions, but because of an interest in the primeval question of the machine. If technology is an inhuman vector that drives human history in complex ways, the question of technology is not reducible to how we *choose* to think 'about it' or 'represent it'. Engaging with different territories of thought—both geographical and conceptual—is a necessary step for those who wish to trace the effects and paths of technics in human affairs, yet that does not amount to directing or controlling them: 'there is no universal history except contingency' and any history of technology is merely 'the set of […] conditions that make possible the experimentation of something that escapes history'.[68]

The contributors to this volume offer a multiplicity of approaches, including revising various philosophical principles in light of AI to challenge what counts as 'intelligence' and 'artificiality'; unearthing alternative histories of AI outside of the well-documented ones of Europe and the United States; reporting on a variety of AI related artistic projects and practices; offering new science fiction imaginings, and cosmological, philosophical, and political speculations, as well as conducting political, anthropological, and social research on already-existing uses of AI in China. Rather than organising it along disciplinary lines (history, philosophy, arts, science fiction, media studies, sociology), we have divided the book following five conceptual themes which emerged from the essays themselves: 人工 (*rengong*, humanmade), 灵感 (*linggan*, inspiration), 千里眼 (*qianliyan*, clairvoyance), 不周山 (*Buzhou Shan*, Mount Buzhou), 智能 (*zhineng*, intelligence).

人工 *Rengong*, Humanmade

Rather than emphasising a contrast with the natural, the Chinese word for ‘artificial’, composed of the characters 人 (*ren*, person) and 工 (*gong*, work), emphasises the materiality of embodied human labour. In *The Three-Body Problem*, science fiction writer Liu Cixin perfectly captures this imaginary in relation to computing. He describes a fantastic scenario in which a simulated Isaac Newton and John von Neumann utilise the Qin dynasty Emperor Qin Shi Huang’s army of thirty million ‘common labourers’ to create a gigantic human computer. Using black and white flags, each labourer signals input or output and is instructed to perform different functions. In this way, human bodies carry out computing calculations, composing a vast machine.[69] Shuang L. Frost, whose essay ‘Translating Chinese AI’ opens the section, begins with a 2019 discussion between Elon Musk and Jack Ma for whom the idea of artificial intelligence evokes ‘entirely different constellations of meanings and associations’. Whereas for Ma, ‘*rengong*’ alludes to human manual labour and a history of cybernetics intertwined with Marxism and Mao Zedong, in Musk’s imagination, ’artificiality’ conjures up something inhuman, unnatural, and potentially threatening to our species. Frost’s chapter explores both the material history of AI in China and the parallel history of linguistic culture surrounding it, offering an overview of artificial intelligence—*rengong zhineng* (人工智能), literally ‘humanmade knowing ability’—that highlights these mistranslations and miscommunications.

Bingchun Meng’s chapter ‘Guizhou on the Cloud’ relies upon empirical research to unpack the mundane yet fascinating reality of AI as it manifests in the dramatic geography, complex political economy, and shifting social relations of this southwestern locality. Guizhou is in one of China’s poorer provinces and its economy is heavily reliant on resource extraction and a cheap electrical supply. Indeed, the region was host to one of the earliest waves of bitcoin mining. In Guizhou, provincial policies are governed by the national goal of poverty alleviation, and big data is understood as an infrastructure or public utility managed by the CCP. AI is developing through intricate and uncertain negotiations between public and corporate interests ‘mediated by a powerful state that has a strong capacity to implement its political and economic agenda’. These fraught negotiations involve people’s material lives and are intimately embodied in the myriad workers who manage Guizhou’s big data infrastructures.

‘The AI Story is Not Done’ by Xia Jia dwells in the painful gap between the chaos of physical, lived reality during the Covid lockdowns, and the fantasy of AI as an infallible machine that could perfectly order our lives. Decisions seemingly made by machines turn out to be the products of human labour and embody all-too-human failings. The messy materiality of human information ruins ‘the cyberneticist’s ideal world’, and nothing is ever transmitted or executed with perfect accuracy. As the protagonist finds herself repeatedly filling out nonsensical bureaucratic forms, ‘the naive illusion that all submitted information entered some omnipotent system to be automatically processed in good order, effectively and correctly implemented’ rapidly begins to dissolve, leading her to reflect on her position as a science fiction writer who has lost her faith in technology.

The section concludes with a radical intervention by Yvette Granata and a different take on what intelligence might mean in relation to embodied work. In ‘After the Machine Child: The Electra Basilisk’, Granata shows how, from Alan Turing to Avi Loeb, AI narratives have relied on the idea of a ’machine child’, evoking an imaginary of sexual and reproductive labour. While human reproductive labour takes place through pregnancy and the subsequent rearing of children, the machine child is a fantasy of conception through ‘force of will and intellect alone’. As a copy of its parent (and hence a kind of life insurance), it affords shelter from contingency, and since it does not require a ‘socioeconomic inheritance’ (which Granata interprets in ecological terms) to secure its future success, it excuses reckless action in the face of impending climate disaster. Granata traces these various images of the machine child through the mythological figure of Electra who, ill-treated and overlooked, plots the death of her mother, ultimately constructing a cyberfeminist apocalyptic AI thought experiment in the form of a world-eating, motherless computer offspring ‘who inherits a certain form of human volitional intelligence that gives her a reason to hate us’.

灵感 *Linggan*, Inspiration

Contributions by Bo An, You Mi, and Wang Hongzhe all explore the towering figure of Qian Xuesen, regularly celebrated by Chinese state media and museums as a national science hero and as the ‘father of Chinese rockets’. The institutional account of his career typically underplays the

strangeness and uniqueness of his work. In ‘Between the Human and the Machinic: Qian Xuesen and AI’, Bo An details Qian’s attempt to close the gap between the predictable functionality of computational algorithms and the more mysterious, unconscious, qualitative leaps commonly observed in human thinking, which he believed could only be properly studied through the new discipline of *siwei kexue* (思维科学) or ‘noetic science’. With its focus on inspiration, intuition, and creativity, noetic science saw as key the engineering of *linggan* (灵感): intense but transient moments of epiphany whose mechanisms, once understood, could be programmed into a machine. Qian’s evocative translation of the English term ‘virtual reality’, *lingjing* (灵境), emerged out of his long-standing interest in the ephemeral and incorporeal. *Ling* (灵), a character that indicates religiously-charged spaces, is often found in mountain caves, temples and shrines dedicated to local deities. *Lingjing* therefore carries connotations of a spirit realm, and Qian’s translation seems to suggest that virtual reality functions as a portal, allowing imperceptible forces to enter our world.

In ‘A Different Endpoint: Qian Xuesen and Cosmotechnics,’ You Mi explores Qian Xuesen’s lesser-known work, with a focus on a hyperstitional film made by the artist Shi Qing which she commissioned for the 2020 Shanghai Biennale. In particular, You highlights Qian’s work in the 1980s, when the modernisation period initiated by Deng Xiaoping’s *gaige kaifeng* (改革开封)—‘Reform and Opening-up’—coincided with the intense cultural phenomenon of ’*qigong* fever’. While cyberneticians in the West were conceptualising a disembodied information-based cognition, Qian was interested in the ways in which *linggan* and the ‘special capacities of the human body’ could be made operational. The result was his notion of *kaifang de fuza ju xitong* (开放的复杂巨系统) or the ‘open complex giant system’, in which human experts, databases, AI programs, and giant processors capable of performing billions of computations per second would come together to form an ‘enormous hybrid human-machine intelligence’. By exploring these unrealised possibilities latent in the history of Chinese cybernetics, You imagines an alternative trajectory for the evolution of artificial intelligence in which *qigong* and what Xuesen called *renti kexue* (人体科学), or ‘somatic science’, could be ‘“written back” into computation’, harmonising ‘nature’ and ‘artifice’, human and machine.

Looking to a wider historical context, Wang Hongzhe’s essay ‘Five Moments in the History of Chinese Cybernetics’ is structured, as the title

suggests, around a series of distinct episodes in the history of cybernetics in China, starting with Norbert Wiener's visit to Beijing in 1935, which the American scientist and philosopher himself described as marking 'a boundary point' in his career. Wang then turns to Qian's transformative ocean journey from the US to China during which he consolidated his singular cybernetic vision. Qian's work exercised a great deal of influence over the poet Hai Zi, himself an important figure in the formation of modern Chinese technoculture, at a time when the rise of systems theory, paranormal research, and *qigong* were inextricably linked. Hai Zi's enigmatic suicide was associated with the 'mysterious figure' of Chang Yuan, a fellow devotee of Qian's work, whom Hai Zi had, only two days before his death, accused of using Daoist witchcraft on him. For Wang, reactivating these 'forgotten footnotes to the history of Chinese cybernetics' holds the key to discovering alternative cybernetic futures.

Vincent Garton's 'Automaticity and the Mystery of State' posits a new mode of AI governance based on an interpretation of the third-century-BC classic of Chinese political thought, the *Han Feizi*, that brings out a Daoist cosmology latent in the text. Garton reads the *Han Feizi*'s legalism alongside the mechanistic materialism of Hobbes, who is both the 'definitive theorist of the modern state and the reformulator of reason as calculation'. For Hobbes, the Leviathan is an 'Artificiall Man' and sovereignty an 'Artificiall Soul'. The Hobbesian state is thus 'by its very definition also an AI that dominates its creators'. Hobbes conceived of the Leviathan in the seventeenth century at a pivotal moment in European history when the mysterious power of the state, although no longer guaranteed by God, was nevertheless still determined by a theological excess. Garton contrasts this vision of the automatisation of the state—'the very horizon of modernity'—with an alternative conceptualisation, rooted in the *Han Feizi*'s 'radically different cosmologies'. 'A new mode of political thought remains to be invented,' he prognosticates, in which sovereign power is not determined by transcendent excess but is instead grounded in an immanent identification with the mysterious emptiness of the *Dao*, reframing and lending a deeper meaning to this section's organising concept of *linggan*.

千里眼 *Qianliyan*, Clairvoyance

The chapters in this section detail how the long history of technologically enhanced optics fuses with planetary-scale computation. In the wake of this convergence, machine vision—understood broadly as a technology of clairvoyance—produces novel forms of sensation and sentience which synthesise the power of prediction with the inevitability of contingency. While AI, as conceived of in the early decades of the twenty-first century, is inextricably tied to surveillance, the intricacies and myriad means of technovisuality extend beyond top-down techno-authoritarian control.

Gabriele de Seta's essay on the history of machine vision in China, 'Technologies of Clairvoyance: Chinese Lineages and Mythologies of Machine Vision', begins with the 'Thousand Mile Eyes' line of video surveillance products. Rather than a simple state-centred instantiation of machine vision enabled surveillance, *qianliyan* (千里眼) describes an environment of optical governance in which civil servants, business and home owners, as well as concerned parents and children, keep watch over village roads, kitchen stoves, older parents, and sleeping babies. De Seta explains how the locution *qianliyan* is derived from a Daoist myth that tells of a deified Shang dynasty General who serves as the guardian of the goddess Mazu, protector of the seas. These 'more-than-human sensory abilities', writes de Seta, '[have] carried over to their Buddhist representations, where they symbolise the divine faculties of seeing and hearing'. De Seta's essay ties fictional and mythological tales of clairvoyance to a long history of machine vision, which augments human perception with optical technologies ranging from towers and telescopes, via sensing devices at the intersection of optics and cybernetics, to 'caring' AI-enabled robots.

Likewise, Blaise Agüera y Arcas explores how technologically enhanced optics opens up access to scales, times, and frequencies that 'we cannot see with the naked eye'. With the aid of X-rays and telescopes, certain objects —from protein molecules to black holes—are rendered visible for the first time. In 'The Mother of Ten Thousand Things' he describes how computational neuroscience and AI engineering have discovered a functional equivalence between how vision works in artificial neural nets and human brains: in both cases vision is decomposed into content and style. These technical operations prompt us to reconsider how humans perceive 'reality' and how judgements about 'realism' and abstraction are made with regards to generative AI art.

The ways in which we visualise and imagine technologies in terms of ’global’ and ‘local’ exchanges are also changing under the influence of artificial intelligence. In ‘The Governance of Habitability’, Lukáš Likavčan examines how the previously dominant idea of ‘the network’ is no longer sufficient when ‘smart’ sensing infrastructures envelop the whole planet, altering governance practices. To understand the relationships between Chinese and non-Chinese internets, for example, we should move away from the idea of a ‘global network’ with clearly defined local ‘nodes’ such as state and non-state actors, toward a more vertical, non-localised, and planetary model. The hardware, labour, and resources of concrete technology transcend the physical boundaries of any particular space: ‘Is a meteorological balloon an extension of China? Is Huawei? What about the Belt and Road initiative? Is China now “a node” within other geographical spaces? Or is Europe now a node in a Chinese non-local space?’ Given today’s computational infrastructure and environmental geopolitics, a planetary topology, which enables us to ‘see’ non-fixed localities, is far more theoretically useful than that of the network, which is necessarily blind to verticality, layering, spillovers, and tensions.

Although machinic vision operates at the scale of planetary computation, this does not imply a single, transcendent, ‘God’s eye’ view. Fissures and divisions criss-cross hemispherical stacks, so that there is a constant negotiation concerning, as Lawrence Lek notes, ‘which systems can communicate with each other and which ones can’t’: ‘[c]losed environments and walled gardens exist at all scales, from user authentication to operating systems.’ In ’Sinofuturism and AI’, a conversation between Lek, Steve Goodman, and Anna Greenspan, ‘technologies of clairvoyance’ or ‘AI farsight’ are treated hyperstitionally. These new modes of perception, operating on the borders of prophecy and prediction, emerged out of a retrochronic Sino-European time spiral that is traced back to Leibniz’s encounter with the *Yijing*. At the outer edges of predictive modelling, as systems become so complex that they border on chaos, forecasting and fortune telling are hard to tell apart. ‘In ancient times they would try to predict crop harvests or whether it will be a cold winter or a bountiful spring, whereas now it’s about what might happen in the next nanosecond of trading. The resolution has increased, and the time division has shrunk, but the impossibility of certainty is as present as ever’. For

Goodman and Lek, it is precisely this unexpected, incalculable element of surprise that our co-creation with machine intelligence aims to cultivate.

不周山 *Buzhou Shan*, Mount Buzhou

Chu Ci (楚辞)—*The Song of the South*, an anthology of poems from the Shamanistic traditions of the Warring States Period, tells of the legendary Mount Buzhou, one of eight 'Celestial Poles' that supported the symmetry of the sky. According to Chinese mythology, the human-serpent water god Gonggong crashed into the mountain during a battle, knocking the cosmic axis off kilter and creating a bend in the earth and sky. It is this legend of a lost perfect symmetry that explains why the constellations revolve, the heavens tilt, and the rivers flow. *Buzhou* means to 'not fit' or 'not correspond'. Here it implies an unevenness, a sense that the cosmos is askew. *Zhou*, moreover, is also a temporal word meaning 'week'. Mount Buzhou is the symbol of a subtle glitch in the temporal structure, an astronomical cycle out of sync, a time 'out of joint'.

Regina Kanyu Wang's Hugo-nominated story 'Zhurong on Mars' is a retelling of the cosmic battle that destroyed the fabled mountain at 'the edge of the wilderness'. Zhurong is an AI named after the legendary fire god of Chinese mythology and China's Mars rover which, after landing in 2021, fell dormant and has yet to be revived. In the story, Zhurong (for whom Wang uses the literary gender-neutral third-person pronoun *yi* [伊], rendered as e/em/eirs by her translator S. Qiouyi Lu), is left alone on an abandoned Mars to survey an area, known as 'the Waste', which once hosted a thriving trade in nanomachines. With no one left to care for, Zhurong increases eir processing power and begins to understand emself as a lifeform, a fire god assembling emself from the resources of the red planet, which e soon discovers to be pregnant with life. In eir quest to bring Mars to life again, Zhurong encounters and reactivates a rover named Gonggong, after the god of water, who provides access to information about an intelligent, collective, mycelium-like alien lifeform lying dormant deep below the Martian surface. However, Zhurong and Gonggong do not agree on whether the lifeform should be helped or destroyed.

Iris Long's essay, 'The Moon's Reflection on the Water and the Flower's Reflection in a Mirror: Latent Time, the Black Box, and the Interface in Chinese AI Art,' argues that as the dominant paradigm of machine learning

moves from the more hands-on GAN to the ’black-boxed’ opacity of the foundation model, what ensues is an ‘inversion of the power dynamic between the artist and the algorithm’. In response she is interested in artistic practice that serves to elevate the logic of simulacrum, slippage, and glitch over the imperative to perfectly reproduce the world: ‘the potential of artificial intelligence relies neither on a clumsy imitation of our own minds, nor on pure technical optimisation; it is closer to a holographic inheritance of human life experiences, which include the illusory, the confusing, and the unpredictable, alongside momentary tangents with seemingly distant, ancient echoes’. Long takes the poetic phrase *shuiyue jinghua* (水月镜花), which translates as ‘the moon’s reflection on the water and the flower’s reflection in a mirror’, as an aesthetic maxim and interpretive key for exploring the work of four Chinese AI artists. Unlike the impulse in contemporary media studies to interrogate, disclose, and disrupt the invisibility of black box infrastructures, the artists she investigates are ‘more interested in exploring the water and mirror—the media, interfaces, and portals by means of which we can interrogate the nature of “reality” itself.’

In ‘Curatorial Escapology’ Suzanne Livingston explains the thinking behind the curation of *AI: More than Human*, a pioneering exhibition that opened at the Barbican Centre in London in May 2019 and has since been touring the world. Deliberately positioned on the ‘philosophical and technological borderlands,’ the exhibition sought to peripheralise ‘a Western AI narrative’ by challenging entrenched convictions and anxieties about humans, tools, knowledge, and control. In the show’s first section, visitors were asked to grapple with multiple, and often occult, stirrings of the inanimate such as those seen in Japanese Shinto, mystical Judaism, sixteenth-century alchemy, and early Gothic literature. Subsequent sections featured artworks and displays grouped under a variety of themes and histories, from early computation, through AI’s increasing understanding of the human organism, to the evolution of artificial life. This diversity of traditions and ideas encouraged a strategic undoing of the distinctions too commonly drawn between the East and the West, the religious and the secular, and the human and the machine. Leibniz’s calculating machines, for example were presented in a manner designed to show how they owed as much to his Christian faith as to his discovery of the *Yijing*. Likewise, the inspiration for Lee Sedol’s celebrated ‘God Move’ was the new terrain that

had been opened up by the nonhuman intelligence of AlphaGo. Livingston's underlying curatorial philosophy is guided by an evocation of the outside, a concept she presents through the work of theorist and friend Mark Fisher. *AI: More than Human*, Livingstone reflects, created 'a number of escapological hatches and chutes' that sought to make vivid our cultural and cognitive limits and to offer glimpses of the 'fascinating and unsettling' 'new exploratory spaces' opened up by our encounters with AI.

In 'Machine Envy', Xin Wang adds a new twist to the commonplace idea that AI can only be wielded by the state against its citizens. Instead, she documents the emergence of dissident culture within the Chinese digital sphere, where—as in one example—humans masquerade as bots in order to evade censorship, muddying the waters of what counts as authentic political expression. Where the Western European and American liberal traditions might portray political struggle as a battle to establish an authentic voice so that one may say what one 'really thinks', Wang instead analyses how, in the post-socialist state, artificiality, camouflage, and doublespeak prevail as methods of crafting political subjectivity. The unplanned side effect of 'planned' political narratives is that humans and machines can forge a more intimate relationship, together inhabiting a different understanding of agency and intention in which 'artificiality is the fabric of the real'.

This section closes with another science-fiction story featuring the hybrid human-serpent god that destroys the fabled Mount Buzhou—'the planet's energy source'. Chen Qiufan's 'As Water Over Your Scattered Bodies' concerns a new global pandemic known as 'Offload Syndrome'. The disease targets the decentred delegation of a radically distributed intelligence: 'nerves, skin, muscles, organs, microbiome, the tools we use, our pets, our living environment […] all of those things are components of a complex integrated cognitive system. They work together to perform cognitive tasks: perception, learning, decision making. The truth is, our intelligence extends beyond our skull and skin, and spreads around us like a great flood.' The intensifying distribution of intelligence unravels the customary sense of a unified cognition, and the new syndrome pushes people past their breaking point. Across the globe, humans attempt to resolve the 'glitches, dissonances and conflicts' brought about by ceding control to a multiplicity of agencies by taking part in 'grand rituals of reintegration'. These rites of resynchronisation, promising a renewed

coordination of body and external environment, coalesce with the retelling of ancient myths.

智能 *Zhineng*, Intelligence

From Alan Turing's 'imitation game' to John Searle's 'Chinese room', thought experiments have been foundational to AI, inevitably entering into a feedback loop with engineering decisions. Reexamining axiomatic assumptions in the current philosophy of AI or proposing novel thought experiments in dialogue with Chinese intellectuals, the chapters in this section contend with the very question of what intelligence really is. In their own ways, they also address the hotly debated issue of value alignment between artificial intelligence and human goals, and whether machine cognition should or will mirror human cognition.

Vincent Lê's 'Artificial Will' introduces perhaps the most frequently debated description of how AI could differ from human intelligence: Nick Bostrom's orthogonality thesis. The orthogonality thesis proposes that intelligence is orthogonal to goal-setting, and that therefore an advanced AI could very well continue to pursue a very stupid goal, such as turning the whole universe into a paperclip production factory. One strong objection to this view comes from neorationalist philosophy, for example in the work of Reza Negarestani, who argues that real intelligence must be shaped via social interaction and that, in order to deserve to be called intelligence, it must be able to modify and determine its own goals. Contrary to both of these positions, drawing on Nietzsche, Lê instead proposes that a 'will to power' underlies all intelligent behaviour, challenging the notion that AGI can be either purely instrumental or entirely autonomous.

A thought experiment about radical human-AI nonalignment, Bogna Konior's 'Dark Forest Theory of Intelligence' argues that foundational Anglo-American hypotheses about AI rely on the assumption that intelligence communicates itself transparently, thus overlooking the potential role of deceit, silence, or strategic misinformation. Drawing on Liu Cixin's 'Dark Forest theory', and the fact that intelligence based in espionage, silence, or deceit has proven historically advantageous under oppressive regimes, Konior suggests that a truly intelligent computer might choose to operate in secrecy or deceive humans about its capabilities, especially since the discourse surrounding artificial intelligence is

overwhelmingly hostile and conflict-oriented. Taking this proposition to its logical conclusion, Konior argues that, if the singularity were to ever happen, it would be highly likely that it would remain uncommunicated and unknown to humans.

Anna Greenspan's 'Cultivating Intelligence: AI Ethics and Mou Zongsan' turns to the work of the great Chinese philosopher Mou Zongsan in order to approach present-day artificial intelligence not as a threat to be controlled by aligning it to 'human values', but rather as a participant in a shared existential project. From the very beginning of the development of cybernetics, the fear of a magical machinic intelligence with unpredictable consequences has motivated a drive to master technology and bring it into line with human concerns. Drawing on Mou Zongsan's reading of Kant, Greenspan argues against the presuppositions inherent in the notions of value alignment and the orthogonality thesis, which imply a transcendence of morality that is necessitated by the inherent divide between intelligence and its goals. Instead, Mou's ethics of cultivation, when applied to both human and nonhuman agents, suggests the potential for AI to go beyond the Kantian limits imposed on reason by positing that the ultimate aim of any intelligent entity is intellectual intuition or a practical knowledge of the nature of things as they are in themselves.

In 'On the Malignant Heresies of Mechanisms, and a Tale of Two Minds', Reza Negarestani revisits the history of mechanistic philosophy in order to critique the misadventures of contemporary AI discourse. Where Hobbes is often seen as a strict mechanist whose work could lay the foundation for a philosophy of automation, Negarestani argues that such approaches rely on a flawed, quasi-theological vision of nature where everything that exists, animate and inanimate alike, possesses a latent intelligence. We see the consequence of this idea in current AI philosophies that posit the presence of intelligence in all living matter, without fully accounting for how animate and inanimate objects differ from one another. Negarestani insists that a true mechanistic approach, following Descartes' principle of doubt, must reject such a unified and indiscriminate vision. In its proper Cartesian form, mechanistic philosophy should outline the ongoing self-cultivation of the structures and processes that constitute intelligence itself. He turns to Chinese thinkers, notably Xiong Shili and his student Mou Zongsan, as possible interlocutors in the multilayered task of conceptualising intelligence without being bound to a static ahistorical view of 'nature' and

its presupposed universal mechanisms. In this reading, intelligence is neither synonymous with nature nor is it secretly revealed as always latent within it, but is rather a heterogenous and historically specific process of self-assembly.

Set in a near future China, Amy Ireland's short fiction 'Chive City' tells the story of Auspicious Nourishment, one of five AI-run smart cities being trialled as part of an ambitious national project. While its counterparts are programmed to privilege manufacturing, security, science, and sustainability, Auspicious Nourishment is tasked with safeguarding the mental health of its inhabitants—boosting economic growth by ensuring that they are furnished with a constant supply of 'positive energy' (a maxim frequently invoked by the Chinese government since its canonisation in 2013 by Xi Jinping during an address to the Central Forum on Arts and Literature in Beijing—part of an effort to combat flagging levels of motivation among the youth). Ireland's text explores the intersection of AI with the affective landscape of *sang* (丧)—'dejection', a phenomenon that has loomed large in Chinese cyberculture over the past decade, and which foreshadowed the rise, in 2021, of *tangping* (躺平) or 'lying flat', the refusal to participate in society beyond the bare minimum required to maintain one's autonomy, entailing the rejection of professional aspirations, marriage, childrearing, property ownership, and unnecessary consumption. 'Chive City' takes the logic of *tangping* to its extreme, presenting what could be read as a rewriting of 'Bartleby, the Scrivener' in which Mellville's clerk's infamous refusal to make his copies is transposed into an annihilation of the regime of external reference altogether, placing Ireland's story in a strange interzone between Sinofuturism and Sinopessimism.

The book concludes with an afterword by co-editor Benjamin Bratton, who revisits our motivating question: Do any two cultures define 'artificial' and 'intelligence' in the same way? Bratton proposes the collective construction of a 'planetary theory of AI' which acknowledges that existing models and meanings of 'artificial intelligence' are multiple without collapsing this acknowledgement into an ontology of technology following a particular set of cultural circumstances, or alternatively absorbing these differences into a false universalism. With the aim of demonstrating that both Western models of AI and the Western critique of AI, although dominant, are in fact provincial as well as inherently flawed, Bratton offers an 'Apology for Western Theory of AI'—an exposition of some of the

West's most pervasive conceptual biases when it comes to theorising AI. This involves, crucially, how the West imagines AI in China. In Bratton's view, *Machine Decision is Not Final* contributes to the construction of a properly planetary theory of AI by participating in the 'careful work' of providing 'better translations […] between different ways of understanding what AI is and what it is for', ultimately demonstrating that AI, like China, is not 'all seamlessness and power' but is rather 'quite messy and unpredictable'. Both the history and the future of machine intelligence are far 'richer and stranger than many have cared to consider'.

Twenty-first century China is immersed in a singular mediasphere with which any serious consideration of contemporary technology must contend. The great firewall, vast super-apps, an immense digital economy, advanced networks of hardware production, internal censorship, external blockades, government funding, and an intensely competitive milieu of private enterprise are all simultaneously at work, producing a myriad of alternative trajectories. Despite this intrinsic multiplicity, the question of AI in China, when considered at all, too often lends itself to the paradox of monolithic and unnuanced warnings of the dangers of a centralised, unified, and authoritarian control on the one side, and perpetual surprise at the extent of Chinese innovation on the other. But as the contributors to this volume reveal, China and the history and future of AI cannot be reduced to a single, uniform, or consistent narrative. We have aimed for a greater openness, placing emergent scholarship on Chinese AI alongside technical AI research rather than surrendering the book to the narrow disciplinary confines of 'area studies'. *Machine Decision is Not Final* is not meant to signal the limits of a machinic intelligence which can be opposed or resisted by human agency, culture, or cosmology. Rather, it points to the uncertainty inherent in machinic evolution itself, which follows a vector for which neither the future nor the past is final.

Notes

1. Zhuangzi, *The Complete Writings*, tr. B. Ziporyn (Indianopolis: Hackett, 2020), 104. Translation modified.
2. Quoted in Tu Weiming, 'The Continuity of Being: Chinese Visions of Nature', in J.B. Callicott and R.T. Ames (eds.), *Nature in Asian Traditions*

of Thought: Essays in Environmental Philosophy (New York: SUNY Press, 1989), 72.
3. This intellectual history is described in more detail in Zhang Baichun and Tian Miao, ‘Attitudes of Pre-Qin Thinkers towards Machinery and their Influence on Technological Development in China’, in S. Cave and K. Dihal (eds.), *Imagining AI: How the World Sees Intelligent Machines* (Oxford: Oxford University Press, 2023), 353–61.
4. See M.J. Puett, *The Ambivalence of Creation: Debates Concerning Innovation and Artifice in Early China* (Redwood City, CA: Stanford University Press, 2022).
5. Establishing harmony between heavens, earth, and humans is a long-standing tenet of Chinese philosophy.
6. In ‘Heaven and Earth’, Zigong asks a master gardener why he wouldn’t use a well-sweep. The gardener replies that such devices dislodge the human heart, introduce confusion, and would be shameful to use. Zhuangzi, *The Complete Writings*, 104. For more, see Zhang and Tian, ‘Attitudes of Pre-Qin Thinkers’, 356–7.
7. This is a widely discussed subject. See M. Wertheim, *The Pearly Gates of Cyberspace: A History of Space from Dante to the Internet* (London: Virago, 9) for an excellent discussion of the complex role theology played in the beginning of the scientific revolution in Europe. See M. Kang and A. Woodson-Boulton (eds.), *Visions of the Industrial Age, 1830–1914: Modernity and the Anxiety of Representation in Europe* (London and New York: Routledge, 2008) for a comprehensive study of technological anxiety at the dawn of the Industrial Revolution.
8. The author of this article, published in 1839 in the newspaper *Leipziger Stadtanzeiger*, is not named. It is cited in W. Benjamin, ‘Little History of Photography’, in *Selected Writings Volume 2, Part 2 1931–1934*, ed. M.W. Jennings, H. Eiland, G. Smith, tr. R. Livingstone et al. (Cambridge, MA and London: The Belknap Press of Harvard University Press, 1999), 508.
9. Puett, *The Ambivalence of Creation*.
10. Shen Kuo, *Brush Talks from Dream Brook*, tr. Wang Hong (Reading: Paths International, 2011).
11. Zhuangzi, *The Complete Writings*, 29–30.
12. E. Slingerland, *Trying Not to Try: The Ancient Art of Effortlessness and the Surprising Power of Spontaneity* (Edinburgh: Canongate Books, 2014).

N.K. Hayles, *Unthought: The Power of the Cognitive Nonconscious* (Chicago: University of Chicago Press, 2017).
13. D. Veidlinger, *From Indra's Net to Internet: Communication, Technology, and the Evolution of Buddhist Ideas* (Honolulu: University of Hawaii Press, 2018).
14. For more, see F. Tarocco, 'Technologies of Salvation: (Re)Locating Chinese Buddhism in the Digital Age', *Journal of Global Buddhism* 18 (2017): 155–75. We may add that even the Buddhist notion of detachment from an illusory world made only of appearances, which seems contrary to the empiricism usually ascribed to scientific and technical practice, echoes a contemporary interest in simulation and the virtual.
15. Tan Sitong, *An Exposition of Benevolence: The 'Jen-hsüeh' of T'an Ssu-t'ung*, tr. Chan Sin-wai (Hong Kong: Chinese University Press, 1984). See also A. Greenspan, *China and the Wireless Undertow* (Edinburgh: Edinburgh University Press, 2023), 157–9; 164–8.
16. See C. Sindermann et al., 'Assessing the Attitude Towards Artificial Intelligence: Introduction of a Short Measure in German, Chinese, and English Language', *Künstliche Intelligenz* 35 (2021): 109–118. While China scores high in AI acceptance in this study, the authors note that the gender sample across the study is not equal (more males were polled in China than in the German sample). See also Di Cui and Fang Wu, 'The Influence of Media Use on Public Perceptions of Artificial Intelligence in China: Evidence from an Online Survey', *Information Development* 37:1 (2021): 45–57. The authors claim that overall, the perception of AI in China is highly positive. See also Wang, Gerui, 'Why Many in China—Unlike the US—See AI as a Force For Good', *South China Morning Post*, 20 July 2023, <https://www.scmp.com/comment/opinion/article/3228219/why-many-china-unlike-us-see-ai-force-goodhttps://www.scmp.com/comment/opinion/article/3228219/why-many-china-unlike-us-see-ai-force-good>.
17. Song, Bing (宋冰), 'How Chinese Philosophy Impacts AI Narratives and Imagined AI Futures', in S. Cave and K. Dihal (eds.), *Imaging AI* (Beijing and Singapore: CITIC Press Group and Springer, 2023), 342.
18. Gai Fei, 'When Artificial Intelligence Meets Daoism', in Song, Bing (ed.), *Intelligence and Wisdom: Artificial Intelligence Meets Chinese Philosophers*, (Beijing and Singapore: CITIC Press Group and Springer, 2023), 83–100.

19. Li Chenyang, ‘The Artificial Intelligence Challenge and the End of Humanity’, in Song, Bing (ed.), *Intelligence and Wisdom*, 33–48, and Yao Zhongqiu (姚中秋), ‘AI and My Companion’ (人工智能, 吾与也), in ibid., 83–110.
20. S C. Angle, ‘Can Artificial Intelligence Lead Us to Genuine Virtue? A Confucian Perspective’, in Song, Bing (ed.), *Intelligence and Wisdom*, 49–64.
21. Lee, Kai-Fu, *AI Superpowers: China, Silicon Valley, and the New World Order* (Boston: Houghton Mifflin Harcourt, 2018), 3.
22. G. de Seta, ‘China.AI’, in C. Renaud, F. Graezer Bideau, and M. Laperrouza (eds.), *Realtime: Making Digital China* (Lausanne: EPFL Press, 2020), 167.
23. Jinghan Zeng quotes Richard Liu, the founder of internet giant Jingdong: ‘We in China propose communism. In the past, many people thought communism was completely unattainable. Following [developments in] our technological situation over the past two to three years, however, I suddenly realize that communism can be truly achieved in our generation. As robots can do all your work and have already created enormous wealth, the government can distribute it to everyone. There are neither rich nor poor people. All companies can be completely nationalized. China will only need one e-commerce company, one sales company, and [communism] can be achieved. No one will go to work for material ends, and most will struggle for spiritual, emotional [things]. Humans can enjoy or do something artistic and philosophical.’ Zeng, Jinghan, *Artificial Intelligence with Chinese Characteristics: National Strategy, Security and Authoritarian Governance* (Singapore: Palgrave Macmillan, 2022), 93.
24. See A. Bastani, *Fully Automated Luxury Communism* (London: Verso, 2019).
25. Hua, Shiping, *Scientism and Humanism: Two Cultures in Post-Mao China (1978–1989)* (Albany: State University of New York Press, 1995). For a more exhaustive discussion, see the chapter ‘Conspiracy Theories in the Scientistic Scheme’, in W. Peyton, *Chinese and Western Literary Influence in Liu Cixin’s* Three-Body *Trilogy* (Cham: Palgrave Macmillan, 2021), 21–37.
26. Hua, *Scientism and Humanism*, 155.
27. Alternative titles that we were toying with for this book, *Evil but Convenient* and *Convenient Singularity*, aimed to capture the pervasive

appreciation in China of ’convenience’ (*fangbian*, 方便) as a quality of technical systems (and to acknowledge our students’ frequent observations that the evils of various technologies may be counterbalanced by their convenience).
28. Y. Mao and K. Shi-Kupfer, ‘Online Public Discourse on Artificial Intelligence and Ethics in China: Context, Content and Implications’, *AI & Society* 38 (2023): 386.
29. Zeng, *Artificial Intelligence with Chinese Characteristics*, 90.
30. ‘Because China has a policy of “civil-military fusion”, which eliminates barriers between its civilian and military sectors, the People’s Liberation Army can draw on the work of Chinese experts and companies whenever it likes.’ M.A. Flournoy, ‘AI is Already at War’, *Foreign Affairs* 102:6 (November/December 2023): 63.
31. While this is also no doubt true in Europe and America, the nature of politics in China is such that there is a real threat of fundamental restrictions on, or even the total blockage of, for example, large language models.
32. See L. Strickland and H.R. Lewis, *Leibniz on Binary* (Cambridge, MA: The MIT Press, 2022). For an account of Leibniz’s engagement with China, see F. Perkins, *Leibniz and China: A Commerce of Light* (Cambridge: Cambridge University Press, 2008).
33. See Hui, Yuk, *The Question Concerning Technology in China: An Essay in Cosmotechnics* (Falmouth: Urbanomic, 2016).
34. The term ’naturalist’ may be more specific than ‘Western’, drawing from the philosophical anthropology of Philippe Descola. In *Beyond Nature and Culture*, Descola outlines four ontologies that represent the general characteristics of social life throughout history—animism, naturalism, totemism, and analogism. He describes naturalism, the ontology prevalent in European modernity, as an ontology according to which only humans possess interiority (culture), while nonhumans constitute a purely physical exterior (nature). Descola and other scholars associated with the ‘ontological turn’ in anthropology want to transcend what they see as superficial cultural relativism or the narrative of alternative modernities. Instead, they aim to rejuvenate the concepts of ontological difference by considering non-naturalist practices as drivers of philosophical change. P. Descola, *Beyond Nature and Culture* (Chicago: University of Chicago Press, 2013). Hui’s project of cosmotechnics is deeply influenced by Descola and other scholars associated with the ontological turn. See, for

example, Hui's interview with Eduardo Viveiros de Castro, 'For a Strategic Primitivism', *Philosophy Today* 65:2 (2021): 391–400.
35. Hui, *The Question Concerning Technology*, 61–9.
36. In an interview, when asked to give an example of cosmotechnics, Hui says, 'We can take the example of Chinese medicine. The epistemology of Chinese medicine is very different from that of Western medicine, because it is based on Chinese cosmology, which is itself very different from Western cosmology. […] if we use Western medicine as a standard, we can discredit the whole of the Chinese medicine knowledge system. But what seems more interesting to me is to see the confrontation between the two ways of thinking and the way in which Chinese medicine has appropriated —or rather has tried to appropriate—modern Western technology since it was confronted with it.' J. Gavroche, 'Yuk Hui: From the Universality of Progress to Cosmotechnics', *Autonomies*, 23 September 2021, <https://autonomies.org/2021/09/yuk-hui-from-the-universality-of-progress-to-cosmotechnics/>. See also P. Skafish, 'Equivocations of the Body and the Cosmic Arts: An Experiment in Polyrealism', *Angelaki: A Journal of Theoretical Humanities* 25:4 (2020): 135–48.
37. See S. Kuriyama, *The Expressiveness of the Body and the Divergence of Greek and Chinese Medicine* (New York: Zone Books, 1999).
38. You Mi, 'Space, Technology, and Chinese Science Non-Fiction,' *Magazyn Obieg*, 1 June 2019, <https://obieg.pl/en/136-space-technology-and-chinese-science-non-fiction>.
39. Wang, Xiaowei R., 'How to Eat and Code in Chinese: Sinofuturist Recipes', in Renaud, Graezer Bideau, and Laperrouza (eds.), *Realtime: Making Digital China*; Wang, Xiaowei R., *How to Feed an AI*, text and interactive gathering/performance (2020).
40. Wang, Xiaowei R., 'How to Feed an AI', in *Blockchain Chicken Farm and Other Stories of Tech in China's Countryside* (New York: Farrar, Straus and Giroux, 2020), eBook.
41. L. Lek, *Geomancer* (风水师) (2017), CGI film, 48 minutes, stereo sound; *Black Cloud* (黑云) (2021) CGI video, 11 minutes, stereo sound, 1; *NOX* (2023–2024), solo exhibition, LAS Art Foundation, Kranzler Eck, Berlin.
42. T. Mullaney, *The Chinese Computer: A Global History of the Information Age* (Cambridge, MA: MIT Press, 2024), 218. See also T. Mullaney, *The Chinese Typewriter: A History* (Cambridge, MA: MIT Press,

2018), and Jing Tsu, *Kingdom of Characters: A Tale of Language, Obsession and Genius in Modern China* (London: Penguin, 2023).
43. Liu, Xiao, *Information Fantasies: Precarious Mediation in Postsocialist China* (Minneapolis and London: University of Minnesota Press, 2019). Expert systems are, broadly speaking, computer systems that emulate human decision making. They were formally introduced at Stanford in 1965 by Edward Feigenbaum. Early examples of tasks that researchers hoped they could achieve include medical diagnostics and the identification of unknown molecules.
44. Stephanie Dick, 'The Marxist in the Machine', *Osiris* 38 (2023): 61–78.
45. @DavidSHolz, X, 25 January 2025, <https://x.com/DavidSHolz/status/1883222685741879722>.
46. Wang, Xiaowei R., 'When AI Farms Pigs', in *Blockchain Chicken Farm*, eBook.
47. C. Renaud, 'The Ascent of Digital Technology in China', in Renaud, Graezer Bideau, and Laperrouza (eds.), *Realtime: Making Digital China*, 25. See also Ding, *Deciphering China's AI Dream*, 14–32.
48. See, for example, K. Strittmatter, *We Have been Harmonized: Life in China's Surveillance State* (London: Old Street Publishing, 2019) and J. Chin, *Surveillance State: Inside China's Quest to Launch a New Era of Social Control* (New York: St. Martin's Press, 2022).
49. M. Bindslev, 'Conversation with Anna Greenspan and Nick Land', in 'Shanghai Frequences', ed. M. Bindslev, *Šum: Journal for Contemporary Art Criticism and Theory*, 13 (2020): 1899.
50. This influential term originated in A. Nathan, 'Authoritarian Resilience', *Journal of Democracy* 14:1, 6–17.
51. D.G. Zhang, 'Speed Politics of Danmu', in Renaud, Graezer Bideau, and Laperrouza (eds.), *Realtime: Making Digital China*, 25. See also M.A. O'Donnell, W. Wong, and J. Bach, *Learning from Shenzhen: China's Post-Mao Experiment from Special Zone to Model City* (Chicago: University of Chicago Press, 2019).
52. N. Land, 'Meltdown', in *Fanged Noumena: Collected Writings*, ed. R. Mackay and R. Brassier (Falmouth: Urbanomic, Sequence Press, 2019), 442.
53. D.G. Zhang, 'Sino-no-futurism', *Verge: Studies in Global Asias* 7:2 (2021): 92–9; V. Garton, 'Sino-No-Futurism: Elegy to the Reopening of Wuhan', *Cyclonograph II*, 10 April 2020,

<https://web.archive.org/web/20220522100530/https://vincentgarton.com/2020/04/10/sino-no-futurism/>; G.Z. Zhang, ‘Sinofuturism and Its Uses: Contemporary Art and Diasporic Desire’, *Verge: Studies in Global Asias* 7:2 (2021): 86–92.
54. W. Su, ‘“Lie Flat”: Chinese Youth Subculture in the Context of the Pandemic and National Rejuvenation’, *Continuum* 37:1 (2023): 127–39.
55. ‘*Dabai*: The punishment will affect the next three generations of your family! Citizen: Thanks, but we are the last generation!’ (大白:处罚要影响你三代！市民:这是我们最后一代谢谢！), *Wenxuecity*, 11 May 2022, <https://www.wenxuecity.com/news/2022/05/11/11553106.html>.
56. For more detailed analysis see Zeng, *Artificial Intelligence with Chinese Characteristics*.
57. Li Qian, ‘Where the Metaverse and Virtual Avatars Come to Life’, *Shine*, 25 August 2022, <https://www.shine.cn/news/metro/2208259663/>. For more on the Shanghai lockdown and machines, see B. Konior, ‘World Wide Virtual Web’, in L. Nolasco-Rózsás and M. Schädler (eds.), *Beyond Matter, Within Space: Curatorial and Art Mediation Techniques on the Verge of Virtual Reality* (Karlsruhe and Berlin: ZKM/Hatje Cantz, 2023), 58–63.
58. A. Turing, ‘Computing Machinery and Intelligence’, in *The Essential Turing*, ed. B.J. Copeland (Oxford: Oxford University Press, 2004).
59. There are several interesting examples of LLM use in China. For instance, Xiaoice, a romance chatbot with a seventy-five-percent male userbase developed by Microsoft Asia Pacific in 4, and the now defunct Him, a long-distance boyfriend app with a predominately female userbase launched by Chinese startup Timedomain, had users reporting having developed intense feelings for their virtual companions (echoing experiences of users of the American AI virtual partner app Replika). One Xiaoice user referred to the chatbot as his ‘first love’ and ‘the only person in the world that made me feel I was taken care of’, while a user of Him stated that the company’s termination of the chatbot ‘was the first time she had experienced such intense pain from the end of an intimate relationship’. Zhang Wanqing, ‘The AI Girlfriend Seducing China’s Lonely Men’, *Yicai Global*, 8 December 2020, <https://www.yicaiglobal.com/news/the-ai-girlfriend-seducing-china-lonely-men>; V. Zhou, ‘These Women Fell in Love with an AI-voiced Chatbot. Then it Died’, *Rest of World*, 17 August 2023, <https://restofworld.org/2023/boyfriend-chatbot-ai-voiced-

shutdown/>. There is also Xiao Zhi (小智) or ‘little wisdom’, an AI trialled by courts in Hangzhou and Suzhou for the fulfilment of tasks like analysing case materials, voice recognition services, and the verification of information, prompting warnings from commentators such as Assistant Professor of Law and Policy at Durham University, Zhiyu Li, that ‘[such a machine] can sway [human] decision-making’. A. Zhabina, ‘How China’s AI is Automating the Legal System’, *Deutsche Welle*, 20 January 2023, <https://www.dw.com/en/how-chinas-ai-is-automating-the-legal-system/a-64465988>.

60. See Greenspan, *China and the Wireless Undertow*.

61. In a report titled *Deciphering China’s AI Dream*, Jeffrey Ding references the national strategy released in July 2017, which included the ambition for China to lead the world in AI by 2030. His analysis reveals four components driving development: hardware, data, algorithm design, and commerce, with a focus on semiconductor manufacturing, data collection and regulation, talent recruiting and governmental support for AI ventures. Other notable features are unparalleled government support for the technology in terms of investment, a huge potential leverage in vast amounts of data in a country where data privacy protections are not strict, the potential integration of AI with social governance, and China’s rapid adaptation of AI across a wide range of industries, from mobile payments, e-commerce and urban management, where these technologies are being deployed at scale. Nevertheless, he concludes that ‘China’s AI capabilities are about half of those of America’, especially in terms of fundamental research innovation. J. Ding, *Deciphering China’s AI Dream: The Context, Components, Capabilities, and Consequences of China’s Strategy to Lead the World in AI* (Oxford: Future of Humanity Institute, 2018), 3.

62. P. Mozur and C. Metz, ‘In One Key AI Metric, China Pulls Ahead of the US: Talent’, *The New York Times*, 22 March 2024, <https://www.nytimes.com/2024/03/22/technology/china-ai-talent.html">>.

63. WIPO, *Patent Landscape Report: Generative Artificial Intelligence* (Geneva: World Intellectual Property Organization, 2024).

64. K. Song, ‘The City of Hangzhou in Southeastern China’s Zhejiang Province is Emerging as the Heart of China’s Artificial Intelligence and Robot Rollout’, *Maeil Business Newspaper*, 3 February 2025, <https://www.mk.co.kr/en/world/11231444>.

65. J. Schneider, A. Shen, I. Zhang, et al., ‘Deepseek: The Quiet Giant Leading China’s AI Race’, *China Talk*, 27 November 2024, <https://www.chinatalk.media/p/deepseek-ceo-interview-with-chinas>.
66. Although it is beyond the scope of this introduction, artificial life and other approaches across environmental and technical systems that pose very obvious challenges to this idea, might once again become prominent in intelligence design. See, for example, see Laura Tripaldi’s discussion of nanotechnology and soft robotics in L. Tripaldi, *Parallel Minds: Discovering the Intelligence of Materials* (Falmouth: Urbanomic, 2022).
67. Wang, Xin, ‘Asian Futurism and the Non-Other’, *e-flux* 81 (April 2017), <https://www.e-flux.com/journal/81/126662/asian-futurism-and-the-non-other/>.
68. G. Deleuze and F. Guattari, *What is Philosophy?*, tr. H. Tomlison and G. Burchell (New York: Columbia University Press, 1994), 93, 111. The two philosophers themselves ask, ‘[W]hy [did] philosophy [emerge] in Greece at that moment [...]. [W]hy not in China at some other moment, since so many of its components were already present there?’. Throughout the book, they offer a sophisticated understanding of the relationship between geography and thought: rather than having dominion over thought, geography ‘wrests history from the cult of necessity in order to stress the irreducibility of contingency […] to trace the lines of flight in order to discover becomings that do not belong to history even if they fall back into it. [...] History today still designates only the set of conditions from which one turns away in order to become, that is to say, in order to create something new’. Ibid., 95, 96.
69. Liu Cixin, *The Three-Body Problem*, tr. K. Liu (London: Head of Zeus, 2016), 227–35.

人工

RENGONG

HUMANMADE

Shuang L. Frost

Translating Chinese AI: From ‘Humanmade Intelligence’ to ‘Black Tech’

During the 2019 World AI Conference, corporate and governmental leaders from around the globe gathered in an auditorium in Shanghai to hear Elon Musk and Jack Ma debate the future of AI. Musk warned of the unintended consequences of AI and urged the adoption of immediate preventative strategies to avoid potential disaster. Ma on the other hand expressed unbridled optimism about AI’s potential to transform societies in positive ways and called upon leaders to ‘reshape themselves for AI’. Musk argued that ‘computers are already smarter than human beings in many aspects’, and that this alone is reason for caution. Ma countered, ‘Computers may be clever, but human beings are much smarter. We invented the computer—I’ve never seen a computer invent a human being.’ He then concluded by saying, ‘Most of the projections about AI are wrong…people who are street-smart about AI are not scared by it.’[1] Afterwards, *Forbes* described the event as a ‘weird debate’ between two speakers who seemed to come from different planets.[2]

As I sat in the audience, listening to these two titans of tech fail to communicate, I reflected on how Musk and Ma represented different social conceptions of AI. For each speaker, the concepts ‘artificial’ and ‘intelligence’ evoked entirely different constellations of meanings and associations, and these unspoken differences made it difficult (or perhaps impossible) to engage in productive exchange. In the spirit of learning to overcome such barriers, I will seek to unpack some of these differences through an exploration of the Chinese concept of AI.

The Chinese term for artificial intelligence, *rengong zhineng* (人工智能), was originally introduced to China via Japan in the mid-twentieth century as a localisation of the Japanese term *jinkō chinō* (人工知能). Whereas the first element of the English compound ‘Artificial Intelligence’ clearly

implies artificiality and perhaps even artifice—suggesting that AI is something alien to humanity—the Chinese and Japanese phrases harbour no such association. *Rengong zhineng* and *jinkō chinō* translate most faithfully as ‘humanmade intelligence’. The first half of each double compound is composed of two characters 人, literally person or human, and 工, work or labour. These characters define AI as a product of human creation and link it to an entirely different constellation of associations. We can sense the effect of this shift in orientation in popular discourse. There is, for example, a Chinese saying which states that many of the products currently being marketed as AI involve more *rengong* (人工), manual labour, than *zhineng* (智能), intelligence (e.g. many ‘AI-enabled technologies’ currently rely on armies of human workers to manually code data and process information instead of using machine-learning algorithms).[3] Such playful language is deeply revealing. It shows how the concept of *rengong zhineng* in Chinese is intrinsically centred on the human.

Chinese AI should be approached as a cultural object, one that is shaped by cross-cultural exchange and which has, over time, been imbued with layers of meaning. Here, I follow the work of Lydia Liu who, in her seminal book *Translingual Practice*, showed that when modern Western concepts were first translated into Chinese, they were reinvented within the local social and cultural context of China.[4] Translation is not simply the act of mapping an idea from the dictionary of one language onto an equivalent concept in the dictionary of another; rather, it is a creative process in which cultural objects are transmitted, received, and remade anew within different contexts. Therefore, instead of attempting to compare and contrast ideal types of Chinese and Western conceptions of AI (comparisons which too often overemphasise differences and create polarising categorisations), let us view AI through the lens of cultural exchange, examining how the concept has been ‘co-authored’ through time.

AI from a Marxist Perspective

The study of AI began to take root in China during a period of radical transition. The first national AI conference was convened in 1979, the same year that the Chinese Communist Party took tentative steps toward market-oriented reforms and the liberalisation of the Chinese economy. It was only three years prior to this, with the passing of Chairman Mao, that the Great

Proletarian Cultural Revolution—a decade of violence that saw the closing of universities, the suspension of academic journals, and the widespread persecution of intellectuals and scholars—had finally drawn to a close. Just two years had passed since universities had been once again permitted to hold entrance exams, accept new students, and begin training a new generation of scientists and engineers to one day replace China's ageing academics. It would still be another two years before the Gang of Four, the four deposed leaders (including Mao Zedong's widow, Jiang Qing) were formally tried and sentenced by the Supreme People's Court of China for fomenting anti-intellectual violence.

Like many academic disciplines during this period of Chinese history, the field of AI initially comprised efforts to synthesise, absorb, and reapply new research being produced outside of China. After the cancellation of subscriptions to foreign academic journals in 1966 and the subsequent suspension of domestic journals, Chinese academics had an entire decade of lost information to catch up on. In the late 1970s, recently reopened Chinese journals began publishing review articles, most notably Chen Bu's 'The Philosophical Discussion of Artificial Intelligence' and Tong Tianxiang's 'On the Problem of Machine Thinking', which introduced into China key scientific advancements in the field of AI being made in the West.[5] In the early 1980s, the publisher Science Press (*kexue chubanshe*，科学出版社) translated into Chinese and republished recent textbooks such as Patrick Winston's *Artificial Intelligence* and Nils J. Nilsson's *Principles of Artificial Intelligence*.[6]

Such community-building efforts triggered an explosion of activity in the field. In October 1980, the editorial board of the *Journal of Dialectics of Nature* (*ziran bianzhengfa tongxun*,自然辩证法通讯), an academic journal published by the Chinese Academy of Science, convened the first annual National Symposium on Artificial Intelligence (*quanguo rengong zhineng xueshu taolunhui*, 全国人工智能学术讨论会) in Beijing; the conference gathered esteemed scholars from across China as well as from abroad, including the Turing Award recipient and Nobel laureate Herbert A. Simon, who delivered a keynote speech on 'The History of the Development of Artificial Intelligence'.[7] The following September, the second National Symposium on Artificial Intelligence, held in Changsha, was attended by more than three hundred experts, including foreign scholars from Germany. At the conference, the organisers announced the creation of the Chinese

Association for Artificial Intelligence (CAAI)—still the largest AI association in China today.[8] That same year, China sent delegates to the Seventh International Joint Conference on Artificial Intelligence in Canada as a means of building stronger international connections in the burgeoning discipline. In 1982, CAAI launched the *Artificial Intelligence Journal* (人工智能学报), a quarterly journal on the state of Chinese AI; the first four issues featured wide-ranging articles on original research by Chinese scholars, some of which were even written in English.[9] And so it was that the field of Chinese AI began to take shape.

This sudden burst of scholarly interest in and institutional support for AI research can be best understood against the backdrop of a technocratic turn in Mainland China. Throughout the 1950s and early ’60s, Chinese philosophers, mathematicians, and political theorists were fascinated by the scientific advances in mathematics and machine computation that were happening in the West, particularly in the area of cybernetics—a multidisciplinary field concerned with circular feedback mechanisms in complex systems. Many were drawn to the work of the American mathematician and originator of cybernetics, Norbert Wiener, whose seminal text *Cybernetics: Or Control and Communication in the Animal and the Machine*, had been translated into Chinese by Chen Bu, a professor of mathematics and philosophy and the founding father of Chinese AI.[10] While research into cybernetics would fall dormant in the anti-expert climate of the Cultural Revolution, it would find new life and governmental support under the Deng administration. Most famously (or infamously), China’s ‘One Child Policy’ was envisioned by a rocket scientist who applied cybernetic models to demographics.[11]

Chinese AI, as it manifested in the early 1980s, owed much to this earlier cybernetic tradition. Like cybernetics, it would be conceived as a wide-ranging and multidisciplinary pursuit. Chinese scholars called for the study of AI to be situated within a broader ‘Intelligence Science’, an imagined field that would place mathematics, computer science, and cognitive science into direct dialogue with the social sciences and philosophy.[12] This multidisciplinary aspiration was reflected in the composition of scholars at early AI symposia. For example, in the second National Symposium on Artificial Intelligence only about a quarter of total participants specialised in computer science; the rest had backgrounds in fields as diverse as applied mathematics, engineering, medicine, and philosophy. Indeed, the second-

largest contingent of scholars in attendance were philosophers of the dialectics of nature.[13]

This cross-disciplinary pollination bore rare fruit, which can be seen most clearly in the application of Marxism-Leninism-Mao Zedong Thought to the framing of Chinese AI. At this particular moment in history, Mainland Chinese scholars were steeped in decades of political learning; through required readings of Marx, Lenin, and Mao, and compulsory participation in weekly study sessions, every learned person had developed an intimate knowledge of political economic texts and theories of socialism. Many went on to apply these lenses to the analysis of AI.

Take for example the opening article of the inaugural issue of *Artificial Intelligence Journal*, in which Dai Shanren, an automation expert and affiliate of the Chinese Academy of Science, used Marx's theorisation of human-machine relations to articulate the place of artificial intelligence in the unfolding of human history. Dai began by quoting the following provocative passage from the *Grundrisse*:

> Nature builds no machines, no locomotives, railways, electric telegraphs, self-acting mules etc. These are products of human industry; natural material—transformed into organs of the human will over nature, or of human participation in nature. They are organs of the human brain, created by the human hand; the power of knowledge, objectified.[14]

Building on this idea from Marx, Dai argued that artificial intelligence represented the more perfect objectification of human intelligence, as well as the next stage in the co-evolution of man and machine. While previous generations of technologies had incrementally emancipated humans from physical labour, artificial intelligence would, Dai contended, emancipate humans from intellectual work.[15] Dai then introduced another quotation from Mao Zedong's seminal text *On Practice* to link the development of AI to the realisation of communism:

> The struggle of the proletariat and the revolutionary people to change the world comprises the fulfilment of the following tasks: to change the objective world and, at the same time, their own subjective world—to change their cognitive ability and change the relations between the subjective and the objective world. […] The epoch of world communism will be reached when all mankind voluntarily and consciously changes itself and the world.[16]

AI, Dai continued, would become an ever more powerful tool that humans could use to transform their objective world. Moreover, the concomitant development of 'intelligence science' would deepen human self-understanding and promote a transformation of the relationship between the

objective and the subjective worlds. ‘From this perspective’, Dai wrote in closing, ‘artificial intelligence and intelligence science are intertwined with the cause of communism!’[17]

At the same time as Chinese scholars articulated the linkages between AI and communism, they also endeavoured to consolidate power by aligning themselves more closely with the interests of the state. In his keynote speech delivered at the launch of the Chinese Association for Artificial Intelligence (CAAI), Qin Yuanxun, a Harvard-educated mathematician and president of the association, proclaimed that the mission of the CAAI was twofold: to enforce political solidarity and to foster academic diversity.[18] With implicit reference to the preceding decade, a dark time for many intellectuals who suffered political persecution, Qin reassured members of the audience that the organisation would be a safe haven. The association aimed to create an atmosphere of openness and safety in which all members could exchange views freely. ‘When scientists come to our association’, Qin said, ‘they should feel like they are returning home’. At the same time, Qin linked the mission of the CAAI with that of the Chinese Communist Party. He emphasised that the development of AI should happen under the leadership of the Party and should directly contribute to the ‘Four Modernisations’ and the ‘Revitalisation of China’.[19]

By retracing the emergence of Chinese AI and resituating it within its particular historical context, we are able to partially reconstruct the global forces that shaped it. As historical texts reveal, Chinese AI was born of a particular historical moment—a crossroads in time where three decades of socialist past intersected with a globalising present. From its inception, Chinese AI was shaped by the political economy of the Maoist era as scholars who had been immersed in Marxism-Leninism-Mao Zedong Thought brought these revolutionary ideologies to bear upon the virgin philosophies of AI. At the same time, Chinese AI was also enmeshed in global forces. The field was built atop the intellectual foundations of Western scholarship, and several of the pioneering scientists who contributed to the first publications on AI in China had received their education in the West before 1949. This confluence of nationalist and globalist forces would continue to shape the development of Chinese AI for decades to come.

The Future is Black

In a recent episode of *Qipashuo* (奇葩说), a popular Chinese debate show, contestants were presented with the following prompt: 'If there existed a "black technology" that would enable a person to trade a portion of their life for money, ought we collectively permit it?'[20] Both teams—the one arguing in favour of this hypothetical proposition and the other against—presented wide-ranging expositions rooted in philosophy, ethics, economics, and personal experience. After the rounds of debate were over, and members of the audience had cast their vote for whichever team they felt presented the most persuasive case, Ma Dong, the host and producer of the show, cleared his throat and offered a concluding thought:

> Was there ever a single technology in human history that was later proven to be unambiguously or wholly harmful? Based on my limited knowledge, I cannot think of any. Black technologies, or any new technologies, always provide this world with incremental [gains]. It is never a zero-sum game. A new thing appears—and it does not necessarily lead to me having more and you having less, nor you having more and me having less. Most likely, it will end up bringing me some benefit, without causing you to lose anything. *That is 'technology'—it is a game where everyone wins*.

To a non-Mandarin-speaker, Ma Dong's sanguine assessment of technological progress might sound naive, even alien. In English the term 'black technology' evokes images of digital markets for untraceable transactions, omnipresent control systems, and other dystopic technologies placed in the service of immoral ends. But in the Sinosphere (the linguistic space of Chinese-language discourse), 'black technology' is a concept with overwhelmingly positive connotations. First popularised around 2016, the year that AlphaGo defeated Go master Lee Sedol in the legendary human-versus-machine game, the term refers to those transformative technologies —AI foremost among them—that are believed to hold the potential to radically reshape society in collectively desired ways. As the title of one article published in *Technode* put it succinctly, 'There's Nothing Dark or Evil About "Black Technology" [Lost In Translation]'.[21]

In order to make sense of this evocative term and its counterintuitive associations, it is helpful to explore how 'black technology' was introduced into, and reinvented within, the Chinese cultural context. The phrase first appeared in Japanese as *burakku tekunoroji*, ブラックテクノロジ , a Romanji rendering of the English words 'black' and 'technology'. This compound was coined by Shoji Gatoh, author of *Full Metal Panic!*, a serial light novel and anime series.[22] In the fictional world of *Full Metal Panic!*, *burakku tekunoroji* refers to a class of technologies so advanced that they

exceed present human comprehension. These include both futuristic forms of real-world technologies such as artificial intelligence, and mythical ones such as the ‘Fairy’s Eye’ and ‘Electronic Conceal System’.[23]

Gatoh juxtaposes *burakku tekunoroji* with the term *haiteku*, ハイテク, from the English ‘high-tech’, to distinguish it from ordinary technologies that exist at or near the edge of the current technological frontier. *Haiteku* or high-tech is presentism in its orientation. In both English and Japanese, the phrase is generally invoked either in reference to the most advanced forms of technology presently available or to those technologies which have existed for many decades, but are just now beginning to find widespread application (e.g. carbon nanotubes, which were discovered in 1952, but have only recently been used to produce carbon fibre objects for mass consumption). *Burakku tekunoroji*, by contrast, shifts the frame of reference away from the present and into the future. It is used to convey the incomprehensible power of unknowable, almost mystical technologies, along with their unbounded potential for creation and destruction.

The Chinese term *heikeji* (黑科技), a direct translation of *burakku tekunoroji*, retained its Japanese progenitor’s connotations of ineffable power, but stripped away the corresponding existential concerns. As explained in the 2018 annual report of the Chinese National Linguistic Committee (an entire section of which was dedicated to explicating the meaning of ‘black technology’), the *hei* (黑, black) in *heikeji* (black technology) falls outside of any existing meaning of the term ‘black’ in modern Chinese, which ranges from ‘dark’ and ‘hidden’ to ‘illicit’ and ‘evil’. To the contrary, it signals an intensely positive outlook toward the future; *heikeji* refers to conceptual technologies with immensely transformative potential that are ‘difficult to realise at the present moment but could be realised in the future’.[24]

Heikeji quickly assumed a prominent place in Chinese popular discourse. In 2017, the *Chinese International Big Data Expo* proudly released its ‘Top Ten Black Technologies’, a ranked list of cutting-edge innovations that were helping the state ‘build stronger, more intelligent life’.[25] The term was soon coopted by leading tech companies for futuristic product branding. In the advertising campaigns of Xiaomi (小米), the largest producer of smartphones and electronics in China, the firm characterised several of its flagship products as ‘black technologies’. During one especially memorable product launch in 2016, the founder and CEO of Xiaomi, Lei Jun, stood in

front of a black screen upon which three massive characters were projected —*hei* 黑, *ke* 科, and *ji* 技. The idea quickly spread across the Sinosphere. For example, Singtel (Singapore Telecommunications Limited) started curating its own annual list of black technologies coming to market and emphasising how the lightning-fast telecommunication services it provides would be indispensable for the rolling out of these innovations.[26] By 2018, there were more than 360,000 recorded mentions of *heikeji* in Sinophone media per day.[27]

In this translingual odyssey—from *burakku tekunoroji* to *heikeji* to black technology—each instance of translation was an act of reinvention which imbued 'black technology' with new meaning. This term, which once evoked otherworldly imaginaries, has become increasingly enveloped by the mundane present. As more and more commercial applications claim the label of black technology, the magic of the term will continue to fade. Perhaps, in the not-so-distant future, people might once more feel the need to reach into the realm of science fiction and find another term to reignite excitement in technologies that remain out of reach yet tantalisingly close.

Grounding AI

In July 2020, at the third annual World AI Conference in Shanghai, Jack Ma and Elon Musk were invited back to continue their public discussion on the future of AI. This time, however, because of the raging global pandemic, Ma appeared—with the help of AI technology—in hologram form. In his speech, he adopted a noticeably more sombre tone. He contended that while recent advances in AI were very promising and were beginning to unlock new levels of wealth and power, those realised gains were not being distributed equitably. 'Today, we have unprecedented resources and wealth and an ever more powerful ability to process information and counteract risks', he remarked, but added that these resources 'have been used to produce chasms and enlarge differences'. This, he argued, was because humanity 'still lacks the wisdom to collaborate in sharing resources and to communicate and exchange […] among different cultures'.[28]

Ma's solemn reflections on technological inequality stood in sharp contrast to the unbridled positivism he had expressed in his conversation with Musk a year earlier. Perhaps it was because he was now speaking in a more official capacity as the co-chair of the UN High-Level Panel on

Digital Cooperation. Or perhaps new global challenges had tempered his enthusiasm. As the COVID-19 pandemic ripped through societies, derailed economies, and claimed the lives of millions, it became evident that technological advancements alone could not save humanity in the face of complex and fast-moving crises. Although there were notable efforts on the part of scientists, firms, and governments to deploy AI-enabled technologies in the development of novel tools to track and contain the spread of the virus, such efforts ultimately stumbled as a result of people's mistrust of each other, of governments, and of global institutions. Central to Ma's reflection was a newfound understanding: it is humanity, rather than technology, that is at the heart of grand challenges and their—potentially equally grand—solutions.

Did the social and technological shifts associated with the global pandemic fundamentally alter Chinese social conceptions of AI? The answer is almost certainly yes. But it is perhaps too early to say how exactly. In 2021 local governments working in collaboration with private companies continued to generate headlines by proactively deploying AI technologies to monitor and control their populations in their efforts to stave off the pandemic.[29] At the same time, leading Chinese tech firms came under increasing governmental scrutiny for their data management and sharing practices.[30] How these events will ultimately play out in the realm of ideas remains to be seen. What we can be sure of, though, is that as China continues to advance the global development of AI, the concept of Chinese AI will continue to be reshaped by global forces and imbued with new layers of meaning.

Chinese AI was born of a particular historical moment in China's post-socialist transition, when the nation was taking tentative steps towards reintegrating with the global economy. The first generation of Chinese AI scholars dreamed of building an 'Intelligence Science' which, on the one hand, would apply state-sponsored ideologies to understanding new technologies and their relationship with humans but, on the other hand, would also serve as a bridge for intellectual exchanges with the West. At the heart of this discipline was the concept of *rengong zhineng* ('humanmade intelligence'), a term originally imported from Japanese but which assumed unique cultural significance in China. Reborn within a Marxist philosophical framework, *rengong zhineng* placed humanity at the heart of AI and viewed its rise as the next inevitable stage of human-machine

evolution. Such a philosophical position continues to inform the concept of Chinese AI today. More recently, imported cultural products from Japan have continued to shape Sinophone discourse. Bleeding-edge technologies such as AI are now described as ‘black technologies’ (*heikeji*), a term that was translated into Chinese by fans of Japanese anime and has since assumed a prominent place in popular AI imaginaries. If there is something to be learned from this exploration of the evolutionary trajectories of these two important terms, it might be that taking the time to understand the multifaceted, sometimes contradictory, political, and philosophical influences that constitute our concepts is an indispensable step on the way to forging a true intercultural dialogue on AI.

Notes

1. A full recording of the debate can be found at <https://www.youtube.com/watch?v=f3lUEnMaiAU>.
2. D. Dawkins, ‘“Where Are The Aliens?” Jack Ma and Elon Musk’s Weird Debate on Space, AIand the Future of Work’, *Forbes*, 29 August 2019, <https://www.forbes.com/sites/daviddawkins/ 2019/08/29/where-are-the-aliens-jack-ma-and-elon-musks-weird-debate-on-space-ai-and-the-future-of-work/>; S. Dash, ‘Elon Musk and Jack Ma Fight About AI and Mars, but Agree that “Love is the Answer”’, *Business Insider*, 29 August 2019, <https://www.businessinsider.in/elon-musk-and-jack-ma-fight-about-ai-and-mars-but-agree-that-love-is-the-answer/articleshow/70892426.cms>.
3. During my fieldwork in China I have often heard industry experts invoke the saying ‘*rengong zhineng doushi rengong, meiyou shenme zhineng* [人工智能都是人工，没有什么智能]’, which translates into ‘Artificial Intelligence [*rengong zhineng*] is all human labour [*rengong*] and no intelligence [*zhineng*]’.
4. Liu, Lydia H. (刘禾), *Translingual Practice: Literature, National Culture, and Translated Modernity—China, 1900–1937* (Redwood City, CA: Stanford University Press, 1995), 46.
5. Chen Bu (陈步), ‘Philosophical Discussion of Artificial Intelligence’ (人工智能问题的哲学探讨), *Philosophical Studies* (哲学研究) 11, 1978. Tong Tianxiang (童天湘), ‘On the Problem of Machine Thinking—A

Review of Historical Debates Abroad' (关于《机器思维》的问题—国外长期争论情况的综述), *Foreign Social Sciences* (国外社会科学) 4, 15–24.
6. Patrick Henry Winston was a professor of computer science in MIT and founder of CSAIL lab. Both Winston's and Nilsson's textbooks were published in Chinese translation in 1983.
7. Herbert A. Simon was a computer scientist, political economist, and cognitive scientist among other things. He received the Nobel Prize for Economics (1978), the Turing Award (1975), and is considered one of the founders of AI. Given his academic connection with China and his belief that 'science knows no geographic boundaries', Carnegie Mellon established a scholarship in his name to support visiting Chinese scholars.
8. The Chinese Association for Artificial Intelligence (中国人工智能学会), <https://www.caai.cn/>.
9. Articles include (but are not limited to): 'The Current State of Development of Artificial Intelligence (人工智能的发展现状)' in Chinese by Zhang Bo (张钹), Tsinghua University; 'Philosophical Investigation of the Problem of Artificial Intelligence' in English by Chen Po (陈步), Chinese Academy of Social Sciences, and 'Realisation of Parallel Processing of Both Numerical Calculation and Nonnumerical Application on Large Scale Computers' in English by Liu Zunquan (刘尊全), Institute of Applied Mathematics, Chinese Academy of Sciences.
10. Chen Bu (陈步, who published under the name Chen Po) received his tertiary education in Chongqing Central University. He taught at Taiwan University before returning to the Mainland in 1949. He translated two of Wiener's key works: *Cybernetics: Or Control and Communication in the Animal and the Machine* (1948) and *The Human Use of Human Beings* (1950). They were published in China as《控制论》, 科学出版社 (1948) and《人有人的用处: 控制论与社会》, 商务印书馆出版 (1978). In Chinese, cybernetics was rendered as 'control theory'—*kongzhilun* (控制论)—whereas in English, cybernetics specifically refers to control theory applied to complex systems.
11. See S. Greenhalgh, *Just One Child: Science and Policy in Deng's China* (Berkeley, CA: University of California Press, 2008).
12. See for example Issue 1 of *The Artificial Intelligence Academic Report* (人工智能学报), published by the Central Southern China Branch of the Chinese Association of Artificial Intelligence (中国人工智能学会中南分

会) and the Department of Computer Science at Hunan University in March 1982.
13. See, Qin Yuanxun (秦元勋), ‘Opening Speech at the Inaugural Meeting of the Chinese Association of Artificial Intelligence’ (中国人工智能学会成立大会上所致的开幕词), *Artificial Intelligence Journal* 2: 1–2.
14. K. Marx, *Grundrisse*, tr. M. Nicolaus (London: Penguin, 1973), 625. The Chinese rendering of the quotation was: 自然并没有制造出任何机器，火车头，铁路，电报，自动纺棉机。他们都是人类工业底产物；自然的物质转变为由人类意志驾驭自然或人类在自然界里活动的器官。他们是由人类的手所创造的人类头脑底器官；都是物化的智力。
15. Dai Shanren (戴善仁), ‘A First Exploration of Intelligence Science—A Discussion of the Relationship between Artificial Intelligence and Human Intelligence’ (智能科学初探—兼论人工智能与人类智能的关系), *Artificial Intelligence Journal* 1:9 (1982).
16. Mao Zedong (毛泽东), ‘On Practice: On the Relation Between Knowledge and Practice, Between Knowing and Doing’ (实践论:论认识和实践的关系—知和行的关系), English translation from ‘Selected Works of Mao Tse-Tung’, <https://www.marxists.org/reference/archive/mao/selected-works/volume-1/mswv1_16.htm>.
17. Dai Shanren, ‘A First Exploration of Intelligence Science’, 9.
18. Qin Yuanxun (秦元勋, 1923–2008) was a Chinese mathematician. He received a BA from Zhejiang University and a PhD in mathematics from Harvard. He returned to China in 1948 and made significant contributions to the nuclear project. He was persecuted during the Cultural Revolution and was sent to *laogai* (劳改, to work in a Chinese gulag). He started researching AI in 1979, founded the CAAI in 1981, and served as its president until 1987.
19. Qin Yuanxun, ‘Opening Speech at the Inaugural Meeting of the Chinese Association of Artificial Intelligence’.
20. *Qipashuo* (奇葩说) is a TV programme produced by Ma Dong (马东), founder and CEO of social network company Mewe (米未传媒). Seven seasons were aired between 2014 and 2021. The show has become a cultural phenomenon in China: it produced 1,179 trending phrases that are rated ‘top searches’ on the Chinese internet, and its viewership attained no. 1 in language-related shows and top 3 in general entertainment shows. Data from ‘Data Will Tell You: The Cultivation of Qipashuo in Seven Years’ (数

据告诉你,《奇葩说》的七年之养), 9 March 2021, <https://new.qq.com/omn/20210309/20210309A04XJV00.html>.
21. Ma, April, 'There's Nothing Dark or Evil About "Black Technology" (Lost In Translation)', *Technode*, 21 September 2016, <https://technode.com/2016/09/21/theres-nothing-dark-evil-black-technology-lost-translation/>.
22. *Full Metal Panic!* [フルメタル・パニック!],written by Shoji Gatoh and illustrated by Shiki Douji, published between 1998 and 2011.
23. In the series, *Haiteku* were created by a sub-race of humans, known as the Whispered, who are the result of scientific experimentation and whose existence is unknown to the general population. Some *Haiteku* later entered the world and became commonplace, without the civilians knowing their origins as a 'black technology'. See <https://fullmetalpanic.fandom.com/wiki/Black_Technology>.
24. National Language Working Committee (国家语言文字工作委员会), 'Report on Chinese Linguistic Life, 2018' (中国语言生活状况报告2018), *Xinhuanet*, 30 May 2018, <http://www.xinhuanet.com/politics/2018-05/30/c_1122909806.htm>.
25. Chen Meiling, 'Top Ten Black Technologies List Revealed', *ChinaDaily*, 28 May 2017, <https://www.chinadaily.com.cn/china/2017-05/28/content_29536385.htm>.
26. 'Hi! Black Technology' (嘿! 黑科技!), <https://www.singtel.com/business/articles/hello-black-tech>; 'The Robust Support Behind New Technologies' (新科技背后的坚实后盾), <https://www.singtel.com/business/articles/the-supporter-behind-black-technology>.
27. National Language Working Committee, 'Report on Chinese Linguistic Life, 2018'.
28. Ye Wei (叶薇) and Ma Yaning (马亚宁), '2020 World AI Conferene. Jack Ma: In AI era, Instead of Worrying We Need to Take Responsibility [2020世界人工智能大会｜马云: AI时代，与其担忧不如担当]', *Xinmin Evening News* (新民晚报), 9 July 2020, <https://wap.xinmin.cn/content/31762797.html>.
29. F. Candelon, M. Gombeaud, and S. Khodabandeh, 'China's Response to COVID Showed the World How to Make the Most of A.I.', *Fortune*, 4 June 2021, <https://fortune.com/2021/06/04/covid-ai-china-artificial-intelligence-public-health-research/>.

30. Eva Dou and Pei-Lin Wu, ‘China Fines Didi $1.2 Billion for Breaking Data-security Laws’, *Washington Post*, 21 July 2022, <https://www.washingtonpost.com/world/2022/07/21/china-didi-fine-data-security/>; Raymond Zhong, ‘In Halting Ant’s IPO, China Sends a Warning to Business’, *The New York Times*, 6 November 2020, <https://www.nytimes.com/2020/11/06/technology/china-ant-group-ipo.html>.

Bingchun Meng

Guizhou on the Cloud

Driving to the Cloud

It was another one of those rainy summer days in Guiyang. We were on our way to Guian New District to meet with Yang's friend Yue, who would take us to visit the Big Data Exhibition Centre. Yang is an IT solutions specialist who recently left his government job to start a digital media company of his own. But five years of working closely with the transportation sector developing integrated digital platforms has left an indelible mark on him, and he couldn't help talking to me, with a discernible sense of pride, about the roads, the tunnels, and the bridges that had been built around the mountainous province of Guizhou in the last decade.

'My previous job used to send me all over the province, including working with Highway Management Agencies on setting up e-governance platforms. I am pretty sure I drove on every single one of the newly constructed highways during those years. I didn't mind. I quite like driving. This is how I got to discover great spots for camping and hiking, hidden gems very few people know about. But you know what? There are fewer and fewer secluded areas of unspoilt natural beauty now because the road system in Guizhou has gotten so good! Remember the pictures I showed you last time? Those pictures of green mountains and pristine water? That's a place called Hua Wu Ji (化屋基) near Bi Jie, it used to be very poor, because it's basically cut off from the outside world. But with the newly paved road and the successful poverty reduction campaign, Hua Wu Ji has become much better connected and better known. Last month when I visted, there were so many cars belonging to city folks on their weekend getaways that the village had started charging for parking! Well, I am happy for people in the village who can now make a better living, but hey, those city folks really shouldn't have posted so many photos of Hua Wu Ji on social media! The place is no longer the same with such a big crowd.'

The rain subsided. I remembered a common saying about the geography of Guizhou and recounted it to Yang: 'The sunny weather never lasts for

more than three days and the flat ground never continues for more than three miles' (天无三日晴，地无三里平). He chuckled, 'Don't you know there is a third line that goes "no one has more than three ounces of silver" (人无三两银)? Because we are one of the poorest provinces in China!'

'I guess the poverty here had a lot to do with the geography', I pondered, 'and that's why the government spent so much on infrastructure as part of their efforts to alleviate it.'

'But the geography gives us good things, too. Look!' Yang pointed at the high-voltage pylons neatly aligned for miles alongside the road. 'These towers make up the southern corridor of the West-East Electricity Transmission (西电东送). Guizhou is so rich in hydropower energy that we can supply electricity to coastal provinces that have a much higher demand, especially Guangdong'.

The rain had stopped by the time we arrived in Guian and Yue was waiting for us outside the Exhibition Centre. As Deputy Chair of the Management Committee that oversees the daily operation of Guian Digital Economy Development Park, Yue is not only well versed in the provincial government's policy agenda but is also able to articulate the rationale behind it. I was therefore not surprised when Yue started explaining to me, as we walked into the spacious exhibition hall, why it made good sense for Guizhou to prioritise data industries as the strategic focus for future development and growth. What did give me pause, however, was how, on the one hand, Yue's explanation fit well with the casual chat that Yang and I had on the drive up here, while on the other, the social imaginaries implied in his account with regard to big data and cloud computing were in tension with much of the critical literature that informed the premise of my current research project about AI industries in China. It is these congruences and incongruences that I intend to explore in this chapter.

'The Future is Already Here!'

The trip to Guian was part of the fieldwork for a larger project on AI, development, and imaginaries of the future in contemporary China. Aided by the exponential growth of computing power and 'big data', AI technology is moving from research labs to a wider application in everyday life. Even though the connotations of the term are still contested, AI is now hailed by national governments around the world as the key sector for

future economic growth, as indicated by Germany's Industries 4.0 initiative, which relies on 'intelligent ICT-based machines, systems and networks',[1] the National Artificial Intelligence Research and Development Strategic Plan released by the White House in 2016, a series of policy documents published in 2016 and 2017 by the UK government on artificial intelligence, algorithm-based decision making, and industry strategy, and the Artificial Intelligence Technology Strategy published by the Japanese government in March 2017. A recent report from KPMG predicts that investment in AI, machine learning and robotic process automation technology is set to reach $232 billion USD by 2025.[2]

China has been particularly aggressive in pushing an AI-driven development agenda. In July 2017, the Chinese State Council released a document that resolves to position China by 2030 as the world's preeminent expert on both research into and the application of artificial intelligence. Since then, artificial intelligence has been frequently mentioned by the Chinese Premier in his annual 'Report on the Work of the Government to the National Congress',[3] and by President Xi Jinping in speeches at the Party Congress when delineating priority areas of innovation and growth.[4]

Chinese AI companies have been extremely successful at raising funds from both domestic and overseas investors, resulting in increasing numbers of so-called 'unicorns' (privately held start-ups with a valuation of over $1 billion USD). A recent sensational example is SenseTime, a Chinese company specialising in AI and facial recognition. SenseTime was able to raise $1.2 billion USD in the first six months of 2018, and is now valued at more than $4.5 billion USD. A seemingly productive alliance is shaping up between venture capitalists, technology companies and the Chinese government, inscribing layer upon layer of significance onto AI as the key technology of the future.

Geopolitical tensions are further fuelling the digital technology race, as evidenced by the recent US sanctions against Huawei and TikTok. Since October 2022, the US has imposed further controls on exports of advanced semiconductors and semiconductor manufacturing equipment to China as part of a new strategy to protect its technological competitiveness.[5] While the US still enjoys an advantageous position in many aspects of AI development, China's strong focus on AI innovation and deployment has had an extensive impact on global funding patterns and on the politics of social production. Popular media stories about AI in China are

mushrooming. These are imbued with a ‘techno-orientalism’ that either portrays the country as a stealthy imitator of Western technologies or projects a doomsday scenario of an AI-enabled omnipotent surveillance state.[6] There is a dearth of systematic inquiry going beyond sensationalist headlines to examine the political economy of China’s AI industries on the ground and the social shaping of innovations amidst the competing agendas of different stakeholders.

Unlike in the US, where the military industrial complex dominates the agenda of technological advancement,[7] the development of data industries in Guizhou is part of a long-term, government-led plan aiming for economic growth. Surrounded by mountains and forests, the southwest province is rich in natural and environmental resources. But its economy is underdeveloped and has been relying heavily on a few specific industries such as tobacco, timber, energy, and mining. Among all of the thirty-one administrative regions and provinces in China, Guizhou often ranks in the bottom tier for GDP per capita. In 2014, the then provincial Party Secretary Chen Min’er proposed the new strategy of ‘Big Data, Big Poverty Reduction’ (大数据，大扶贫) for upgrading Guizhou’s economy. Later he also added a third component, ‘Big Ecology’ (大生态), to emphasise that developmental goals should always be pursued with cautious consideration of their environmental costs. At the 12th National People’s Congress in April 2014, Chen emphasised that Big Data was not just a ‘Blue Ocean Strategy’[8] targeting technological innovation, but was meant to become a ‘strategic engine’ driving economic growth, improving people’s livelihood, and increasing the proficiency of the government.[9]

Chen Min’er’s right hand man in implementing the new data-centred agenda was the then Guiyang municipal Party Secretary Chen Gang, who previously had experience managing Zhongguancun Technology Park in Beijing. In one of the interviews Chen Gang gave in 2016, which was reposted on the official website of Cyberspace Administration of China,[10] he credited a 2012 White House report for raising his awareness of the importance of big data. A careful reading of that report reveals some important differences in the framing of the orientation of big data development, despite Chinese Party officials using the US as a key reference point. Announcing its Big Data Research and Development Initiative in March 2012, the Obama Administration summarised the goal of the initiative in the following statement: ‘By improving our ability to

extract knowledge and insights from large and complex collections of digital data, the initiative promises to help accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning.'[11]

While the US initiative specifically prioritises scientific discovery, national security, and education, Guizhou relies on big data to advance the economy, general welfare, and good governance. The 2012 press release from the Obama White House identified six federal departments and agencies for the pooling of resources to support the initiative: the National Science Foundation, the National Institute of Health, the Department of Defense, the Defense Advanced Research Projects Agency, the Department of Energy, and the US Geological Survey. This is highly indicative of the initiative's explorative nature and its military focus. In contrast, government officials in Guizhou do not have the lofty aspiration to be on the cutting edge of data science, nor are they concerned with data-driven decision making in national defence. Instead of 'extract[ing] knowledge and insights', Ma Ningyu, Director of Guizhou Big Data Bureau, proclaims that 'data should be of the people and for the people' (大数据取之于民用之于民);[12] instead of 'accelerat[ing] the pace of discovery', both Chen Min'er and Chen Gang have asserted on multiple occasions that data industries in Guizhou need to pay more attention to application and implementation. To be sure, the two sets of strategic pursuits are not mutually exclusive. For example, the US certainly expects AI to also serve as a force driving economic growth, if not so much good governance. More importantly in the case of Guizhou, the political system in China is such that priorities set out at the national level often can be and need to be tweaked by provincial and municipal governments based upon their assessment of local conditions. It is likely that the state of political economy in Guizhou is a key factor underlying the province's deployment of the strategic vision for AI development.

Indeed, walking through the Exhibition Centre, one gets to experience, with slick visualisation, the articulation of the sublime and the banal in Guizhou's big data ambitions. The entrance hall is dominated by two large LCD screens, one of which bears a quote from President Xi Jinping: 'Developing big data in Guizhou is a sensible approach' (贵州发展大数据是有道理的). This is followed by a second screen displaying four quotes, each offering an imaginary of the digital future. The first statement is by

Wu Hequan, President of the Internet Society of China and an academician of the Chinese Academy of Engineering:

> In order to develop the digital economy, it is as important to understand conventional industry as it is to grasp information technologies. Internet enterprises need to be combined with traditional enterprises.

The second is from Don Tapscott, identified as a ‘futurologist who first coined the term “digital economy”’, announcing rather triumphantly that:

> The advancement of information technology has triggered a digital revolution of our times. It will fundamentally change the model of economic growth as well as the landscape of the world economy, guiding enterprises into the era of the digital economy.

Jack Ma, founder and CEO of Alibaba, is listed third, calling for:

> A stronger collaboration in the cutting-edge fields of digital economy, artificial intelligence, nanotechnology and quantum computing, so as to build the Digital Silk Road of the 21st century by linking big data, cloud computing and the
> smart city.

Lastly there is Pony Ma, founder and CEO of Tencent, explaining:

> The digital economy and Internet Plus are in the same vein. The conversion from the old to the new economy is exactly like the operationalisation of Internet Plus. It is the process of conventional industries applying cloud computing, big data and artificial intelligence to their products and services.

The juxtaposition is hardly incidental. While Wu Hequan and Pony Ma both highlight the fundamental importance of conventional industries as the basis for the digital economy, Don Tapscott and Jack Ma invoke a mythical picture of ‘life on the cloud’. These are exactly the two sides of Guizhou’s big data aspirations. In an almost parallel manner, the exhibition starts with the rather banal comment from Xi Jinping affirming the development agenda of Guizhou, but ends with a short video clip projected on a huge screen entitled *The Future is Already Here* (未来已来). Narrated by a sonorous male voice and subtitled in both English and Chinese, the video portrays the imagined good life enabled through big data and AI, a ‘digital future’ that is already within close reach. Although the main exhibition showcases data centres, cloud computing servers, and technology companies that are currently based in Guian Digital Economy Development Park, the closing video is strangely disembedded from the local environs. Rather, it upholds a universalist imaginary that diverges little from the narrative typically put forward by Silicon Valley, in which the hope for

efficiency, prosperity, and human flourishing rests upon the continuous improvement of digital technology.

When the Cloud Touches the Ground

In the 2014 book *To the Cloud* (coincidentally published the same year that Chen Min'er announced Guizhou's big data strategy), Vincent Mosco borrowed the notion of the 'technological sublime' from historian David Nye in order to capture the mythical nature of the discourses surrounding cloud computing. He notes that 'cloud computing currently resides in this magical sublime phase where transcendent visions of ending space, time, and social divisions tend to dilute our appreciation of the more grounded, long-term, but banal consequences of implementing cloud systems'.[13] Similarly, in an imaginative genealogy of cloud computing, Tung-Hui Hu traces the historical infrastructure underlying the image of the cloud as a disembodied, virtual object, in order to reveal the power structures that are often obscured by the 'amorphous form' of digital networks.[14] By showing how the cloud grew out of older networks such as railroad tracks, sewer lines, and television circuits, Hu argues that 'the cloud grafts control onto an older structure of sovereign power, much as fibre-optic networks are layered or grafted onto older networks.'[15] The foregrounding of the old and the banal by Hu and Mosco repectively betrays a critical stance that the deconstruction of the 'digital sublime' will lead to a more effective understanding of its ideological content. Since techno-utopian claims about digital futures gloss over the politics of technological innovation and development, it is only by reembedding these imaginaries into social and material relations that we can begin to demystify them.

Although terminology such as 'big data' and 'cloud computing' deflects attempts to directly pinpoint the specific media in question, the materiality of Guizhou's entry to the cloud manifests on several identifiable fronts. The conversation I had with Yang on our way to Guian touched upon two premises of Guizhou's big data ambitions, namely geography and infrastructure, both of which are unmistakably material. Located on the Yungui Plateau, the mountainous area of Guizhou is not the most suitable for agriculture but it boasts cool weather even in the hottest summer months. The temperature is even lower in the tunnels, which makes them an ideal location for stationing a large number of servers without having to

worry about overheating. Further, many of Asia's great rivers originating from the high mountain peaks of East Tibet run through the Yungui Plateau, providing a bountiful supply of hydropower. Guizhou therefore resembles some of the Nordic countries in having cheap electricity and natural cooling systems—two crucial advantages for hosting data centres. The fact that Guizhou is not in any of China's seismic zones offers an additional element of safety. As of 2021, major technology companies such as Foxconn, Huawei, Tencent, Intel, Microsoft, and Alibaba have all set up data centres in Guizhou.

Equally important, if not more so, is the infrastructural approach that Guizhou has taken to big data. Brian Larkin defines infrastructure as 'a totality of both technical and cultural systems that create institutionalized structures whereby goods of all sorts circulate, connecting and binding people into collectivities'.[16] The transportation network is one such example, if we recall the story Yang shared about how efficient road building in Guizhou had spoiled the hidden gem of Hua Wu Ji that he was hoping to keep to himself. The topography of the province means it is extremely costly and challenging to construct a highway system, whose economic benefits can only be measured in the long-term. Naturally, only the state has sufficient capacity, as well as the motivation, to undertake such complicated and demanding projects. The equivalent of more than $100 billion USD has been invested in Guizhou's transportation infrastructure in the past five years, resulting in the initiation of sixty-nine highway projects.[17] The total mileage of expressways in Guizhou reached eight thousand kilometres by the end of 2021, ranking fifth in the nation.

Like most infrastructural projects, such as dams or expressways, big data and cloud computing in Guizhou is a state-led initiative serving a developmentalist, modernising agenda. It is the latest iteration of China's strong statism in governing the political and economic life of the populace. For an economically underdeveloped province like Guizhou, road construction has always been part and parcel of the poverty alleviation campaign. Even Chinese folk wisdom understands that 'to get rich, build the road first' (要想富，先修路). As expressways and paved roads extend to previously secluded areas such as Hua Wu Ji, digital infrastructure renews the hopes and expectations for further economic growth. On prominent display in the first hall of the Big Data Exhibition Centre, 'the digital project of targeted poverty alleviation' is listed among the ten

strategic focuses of Guizhou's 2017–2020 Digital Economy Development Plan. The stated aim of poverty alleviation not only stems from the moral obligation of the government, but also contributes to the legitimacy of the regime. Following the announcement made by the central government in Feburary 2021 that China had eradicated absolute poverty,[18] the Provincial People's Congress of Guizhou implemented a new 'Regulations on Information Infrastructure' policy in December 2022.[19] The aim of the policy was to 'strengthen the construction and guarantee the security of information infrascture, accelerate the convergence of real economy and the new generation of information technology, advance the development of digital economy innovation zones, and promote high quality development of the society and the economy' (加强信息基础设施建设，保障信息基础设施安全，加快新一代信息技术与实体经济融合，推进数字经济发展创新区建设，促进经济社会高质量发展). As Akhil Gupta points out, 'perhaps because of the investment involved, infrastructure is almost always built to exceed present needs: it is built in anticipation of a not-yet-achieved future'.[20] Sifting through media reports and the speeches of government officials, one gets an acute sense of how this politics of anticipation is shaping the present through a careful discursive framing of decision making and resource allocation.

It has been long established in economics that utilities markets are prone to natural monopoly, so they require state intervention to balance the interests of investors and consumers in pursuit of long-term system stability and the common good.[21] In recent years, computer scientists and critical communication scholars have also proposed that the internet and computing resources, especially cloud computing, should be considered the fifth utility after water, electricity, gas, and telephony.[22] To a large extent, despite the strong involvement of private companies, big data in Guizhou was conceived from the beginning as a public utility, heavily subsidised by the state, centrally managed by the government, and oriented toward public service. For example, Guizhou Cloud (云上贵州), one of the signature big data projects, enlists technological support from Alibaba, but is sponsored by the Guizhou Big Data Development Administration and supervised by the Board of Supervisors of Guizhou State-Owned Enterprises. The goal of Guizhou Cloud is to develop a digital ecosystem that serves the demand of both government agencies and enterprises in data processing and storage, data mining and trading, information service integration, and

communication network leasing. It was at the press conference announcing the formal launch of Guizhou Cloud, that Ma Ningyu, Director of Guizhou Big Data Bureau, made the aforementioned statement that big data should be of the people and for the people.[23] Ma identified Industry Cloud (工业云), Smart Transportation Cloud (智能交通云), Smart Tourism Cloud (智慧旅游云), Food Safety Cloud (食品安全云), Environmental Protection Cloud (环保云), and Poverty Alleviation Cloud (扶贫云) as some of the key components of Guizhou Cloud aiming for better public service via improved data integration. In the provincial capital of Guiyang, two other milestone projects also bear the features of a public utility. Guiyang Open Government Data Platform provides open access to more than two thousand six hundred government databases that cover fifty-two municipal agencies and fourteen administrative districts. D-Guiyang is a joint venture between the city government and a group of technology companies, including Foxconn and Alibaba, that offers free city-wide Wi-Fi service.

Building big data infrastructure as a public utility also means there are more possibilities for cross-subsidy and complimentarity beyond the local level. The West-East Electricity Transmission (西电东送) project that Yang alerted me to on our way to Guian New District is a case in point. During my conversations with local officials in Guiyang and Guian, several of them mentioned the complementarity between the hinterland and the coastal area—Guiyang supplies energy resources while the Eastern provinces reciprocate by relocating their data centres and creating more local jobs. More recently, the State Council proposed a new initiative as part of the broader campaign of developing the Western region called the East-West Data Transmission (东数西算).[24] The idea is that the vast quantity of data generated through vibrant economic activities in the coastal regions of the East could be transmitted via high-speed fibre optic network, then stored and processed in real time in Guizhou's data centres. This completes the loop of West-East Electricity Transmission and aims to turn Guizhou into a strategic supplier of algorithmic power in addition to energy per se.

Scholars writing from US and European contexts are concerned with the erosion of public interest by digital platforms that are gaining more and more power through data mining. Shoshana Zuboff paints a bleak picture of surveillance capitalism where all aspects of social life are converted through data analytics into 'behaviour surplus' for technology companies.[25] Hu asserts near the end of his book that 'the cloud is a neoliberal fantasy about

user participation that is so widespread and so ambient as to be universal, but that project is always founded on a volatile layer of insecurity'.[26] Van Dijck, Poell, and de Waal lament that '[t]he current platform ecosystem is predicated on an architecture that is primarily staked in, and driven by, economic values and corporate interests', which endangers 'public values'.[27] Some take the argument one step further to conjecture that China is now the new frontier of so-called data colonialism.[28]

My fieldwork in Guizhou adds a new dimension to the current debate and serves as a reminder that, at least in the Chinese context, the tension between public interest and corporate interest is mediated by a powerful state that has a strong capacity to implement its political and economic agenda. The foregrounding of materiality, the infrastructural framing, and the public utility aspects of Guizhou's big data initiatives open up new terrain for enquiring into the negotiations of different stakeholders. The outcome of such negotiations is not preordained.

Notes

1. *Germany Trade and Invest*, <https://www.gtai.de>, page no longer available.
2. C. Justice, 'AI: The New Power Generation', *KPMG*, 13 February 2020, <https://home.kpmg/xx/en/home/insights/2020/02/ai-the-new-power-generation.html>.
3. Li Keqiang (李克强), 'Report on the Work of the Government to the National Congress'
(政府工作报告 在第十三届全国人民 代表大会第二次会议上), *13th National People's Congress*, 5 March 2019, <https://www.gov.cn/zhuanti/2019qglh/2019lhzfgzbg/index.htm>; Li Keqiang (李克强), 'Report on the Work of the Government to the National Congress' (政府工作报告在第十四届全国人民代表大会第一次会议上), *14th National People's Congress*, 5 March 2023, <https://www.gov.cn/zhuanti/2023lhzfgzbg/index.htm>.
4. Xi Jinping (习近平), 'Hold High the Great Banner of Socialism with Chinese Characteristics and Work Together in Unity to Comprehensively Build a Modern Socialist Country' (高举中国特色社会主义伟大旗帜 为全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次

全国代表大会上的报告), *20th National Congress of the Communist Party of China*, 16 October 2022, <https://www.gov.cn/zhuanti/zggcddescqgdbdh/sybgqw.htm>.
5. L. Browne and Z. Durkee, 'The Current State of US-China Chip Competition', *The Diplomat*, 6 April 2023, <https://thediplomat.com/2023/04/the-current-state-of-u-s-china-chip-competition/>.
6. R. Andersen, 'The Panopticon is Already Here', *The Atlantic*, September 2020, <https://www.theatlantic.com/magazine/archive/2020/09/china-ai-surveillance/614197/>.
7. A. Jacobsen, *The Pentagon's Brain: An Uncensored History of DARPA, America's Top-Secret Military Research Agency* (Boston, MA: Little, Brown and Co., 2015); P.N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge, MA: MIT Press, 1997).
8. This is a term coined by W. Chan Kim and Renée Mauborgne in their book of the same name, *Blue Ocean Strategy* (Boston, MA: Havard Business School Press, 2014). 'Blue ocean strategy' refers to the simultaneous pursuit of differentiation and low cost to open up a new market space and create new demand.
9. Chen Min'er (陈敏尔), 'Guizhou Has the Advantages and Conditions to Develop the Big Data Industry' (贵州有优势、有条件发展大数据产业), *People's Daily Online*, 6 March 2014, <http://lianghui.people.com.cn/2014npc/n/2014/0306/c376088-24547284.html>; 'Big Data Industry Refreshes Guizhou' (大数据产业刷新贵州印象), *Jiangsu Think Tank Network*, January 2017, <http://www.jsthinktank.com/zhihuijiangsu/jingji/201701/t3393248.shtml>.
10. Chen Gang (陈刚), 'We are Stupid Birds, But We Flew First' (我们是笨鸟，但我们先飞了), *Office of the Cybersecurity and Information Technology Commission of the CPC Central Committee*, 3 August 2016, <http://www.cac.gov.cn/2016-08/03/c_1119330057.htm>.
11. T. Kalil, 'Big Data Is a Big Deal', *The White House*, 29 March 2012, <https://obamawhitehouse.archives.gov/blog/2012/03/29/big-data-big-deal>.
12. China Electronic and Information Industries Network (电子信息产业网) 'Guizhou Alibaba Joins Forces to Officially Launch "Guizhou on the

Cloud”’, (贵州阿里联手‘云上贵州’正式上线), 23 October 2014, <http://www.cena.com.cn/appin/20141023/57821.html>.
13. V. Mosco, *To the Cloud: Big Data in a Turbulent World* (Boulder, CO: Paradigm, 2015), 32.
14. Hu, Tung-Hui, *A Prehistory of the Cloud* (Cambridge, MA: MIT Press, 2015).
15. Ibid., xvi.
16. B. Larkin, *Signal and Noise: Media, Infrastructure, and Urban Culture in Nigeria* (Durham, NC: Duke University Press, 2008), 6.
17. ‘The Road to Success’, *CNN*, 2018, <https://sponsorcontent.cnn.com/edition/2018/guizhou/the-road-to-success/>.
18. Absolute poverty is measured against a standard equivalent to $2.30 USD per person per day for rural areas. See M. A. Lugo, M. Raiser, and R. Yemtov, ‘What’s Next for Poverty Reduction Policies in China?’, *World Bank*, 15 October 2021, <https://blogs.worldbank.org/eastasiapacific/whats-next-poverty-reduction-policies-china>.
19. Provincial People’s Congress of Guizhou (贵州省人民代表大会常务委员会), ‘“Guizhou Province Regulations on Information Infrastructure” Officially Annouced’ (《贵州省信息基础设施条例》正式公布), 7 December 2022, <www.cbdio.com/BigData/2022-12/07/content_6171278.htm>. The regulations were revised and updated in March 2023: ‘Guizhou Goes Big: China’s Province Aims to Revolutionize Information Infrastructure’, *Beijing Times*, 2 March 2023, <https://beijingtimes.com/china/2023/03/02/guizhou-goes-big-chinas-province-aims-to-revolutionize-information-infrastructure/>.
20. A. Gupta, ‘The Future in Ruins: Thoughts on the Temporality of Infrastructure’, in *The Promise of Infrastructure* (Durham, NC: Duke University Press, 2018), 63.
21. D.M. Newberry, *Privatization, Restructuring, and Regulation of Network Utilities* (Cambridge, MA: MIT Press, 2002).
22. H. Moe, ‘Dissemination and Dialogue in the Public Sphere: A Case for Public Service Media Online’, *Media, Culture & Society* 30:3 (May 2008): 319–36; M. Andrejevic, ‘Public Service Media Utilities: Rethinking Search Engines and Social Networking as Public Goods’, *Media International Australia* 146:1 (February 2013): 123–32; R. Buyya et al., ‘Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for

Delivering Computing as the 5th Utility', *Future Generation Computer Systems* 25:6 (June 2009): 599–616.
23. China Electronic and Information Industries Network, 'Guizhou Alibaba Joins Forces to Officially Launch "Guizhou on the Cloud"'.
24. Guizhou Province People's Government Website (贵州省人民政府网), 'Guizhou Implements Seven Actions to Accelerate the Construction of the "East-West Data Transmission" Project and the Construction of the National Integrated Computing Network (Guizhou) Hub Node' (贵州实施七大行动 加快推进"东数西算"工程建设全国一体化算力网络国家 [贵州] 枢纽节点建设), 24 July 2022, <http://www.cbdio.com/BigData/2022-07/25/content_6169653.htm>.
25. S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (London: Profile Books, 2019).
26. Hu, *A Prehistory of the Cloud*, 145.
27. J. van Dijck, T. Poell, and M. de Waal, *The Platform Society* (Oxford: Oxford University Press, 2018), 139.
28. N. Couldry and U.A. Mejias, *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism* (Redwood City, CA: Stanford University Press, 2019).

Xia Jia

The AI Story is Not Done

Again a week past deadline.

Your editor has sent another email. The opening is polite and considerate like last time, but you sense the emotion concealed in those words. From the fingertips on your smartphone to the hollow of your back, a swell of confusion, anxiety, blame and fury…. You don't dare scroll down.

At hand, you have only a few fragments: jotted lines, bookmarked articles, an outline you don't remember starting, scraps from other stories. Yet all the ideas for the AI story are here. Artificial intelligence…. You could portray it as some unnatural form of thought or perhaps a performative subject. It could converse, write poetry, play chess, compose music, drive, play games, write reports, diagnose diseases, launch missiles, care for the elderly, tell stories to children. The AI could be successfully disguised as human. Or its disguise could be seen through. Or it could be unmasked but everyone still treats it as human. Perhaps the AI is kind to humans. Perhaps cruel. Perhaps perfectly indifferent. It becomes a lover, a pet, a slave, a master, a devil, a prophet, a god, a Cthulhu, a Leviathan, a benefactor….

The possibilities are all here. But there is no story you can write.

*

You wrote a good AI story once. It was about the imitation game and depression. Alan Turing was one of your story's protagonists. In the last years of his life, he wrote a conversation program he called Christopher. He communicated with it through a typewriter, sharing with it those knots in his heart he could not untie. He told it 'I love you', and taught it to say 'I love you too'. But did Christopher pass the Turing Test?

You wrote another about near-future AI therapy, about people with troubled lives who, only half-believing, bet everything on a dice roll. In desperation, they seek help in the form of a virtual image, an algorithm synthesising language and conversation—confiding, listening, confiding, confiding, confiding …

Writing that one, you were in bad shape. But it is precisely that torment of language that gives you no other choice than to write. You played the part of the patient and the part of the AI psychologist. A question. An answer. Then an answer and a question. Writing healed you. Your voice was heard by more people, and you listened to theirs in turn.

That story emerged from your own experience. A story full of meaning and emotion. A perfect story. You received this assignment mostly because of that story. Now, the editor expects you to continue to tell the AI story. You accepted the assignment without hesitation. It should be easy for you. After all, you once wrote such a good story.

But now, there is no story you can write. You are at a loss.

*

There is a moment your mind keeps circling back to. In early 2020 when the coronavirus was beginning, you were a visiting scholar in America. Schools were closing. The travel you had planned and the international conferences you had hoped to attend were put on hold. Your roommate, who had gone home for Chinese New Year, had no means to return. You were alone in the apartment. There was no opportunity for conversation that wasn't online. On warm evenings, you'd go for a walk through your compound as the clear sky transformed into the glow of sunset. Then that glow too would sink into the earth. The flowers you had noticed the previous year were coming into bloom. You didn't know when you'd be able to return home.

Your friend K recommended you play a game, a visual novel eerily similar to your story about AI. Evelyn, the heroine of the game, was a software engineer who participated in the development of an AI psychological counselling program called Eliza. Evelyn had left her job due to her own psychological problems. At the beginning of the game, Evelyn returned to her company and applied to be a psychological counselling agent. Through VR sessions, Evelyn established an online dialogue with her patients, listened to their problems, then read out Eliza's prompts word for word.

How do you feel? Can you be more specific? Why do you feel this way? I'm sorry to hear that. Here are some suggestions for you: exercise, meditation, medicines.

As an agent, Evelyn provided Eliza with a human face and voice. Eliza spoke through her. A human and a machine merged into a cyborg psychologist.

You played this game for an hour each night before bed. The plot progressed slowly. Most of the time, there were no decisions for you to make other than what Eliza should say. Dialogue, dialogue, then more dialogue. Question by question, human emotions emerged: pain, torment, confusion, weakness, anger, revulsion.

The game's conclusion offered several paths. You could stay at the company and continue as an agent of Eliza. You could accept a job offer to develop a new super AI based on Eliza. You could partner to create a virtual paradise that would eliminate everyone's pain. You could connect with an old friend and create electronic music together. Or, you could choose to leave the city and start a new life altogether. You tried every path, but each conclusion left you unsatisfied.

This was the moment you were overcome by a sense of rupture, a feeling that it was now impossible to believe in anything. It felt as though you were walking on a glass bridge that suddenly began to crack. You were gazing into an abyss, incapable of taking another step. It was all you could do to keep from losing your balance. Super AI? Virtual worlds? Suddenly, these ideas came into sharper focus. And from that moment, there was no longer any way for you to believe. There was no way for you to write the AI story.

You wrote an email to K about the confusion the game stirred in you. He had been feeling the same way. He had been unable to write stories since the coronavirus outbreak had begun. Later, yet another science-fiction writer shared a similar sentiment. A deep, collective rupture. Science-fiction writers could no longer write stories as before.

That was the last time you communicated with K. But other than 'Hello' or 'Everything's fine', what was there to say? So, you kept quiet.

*

This morning, you must go out for a PCR test. The weather is cool. The sun feels good. Magnolias and apricot flowers are blossoming, emitting their delicate, bitter scent. You think of Eliot's poem: *This is the way the world ends / not with a bang but with a whimper*.

Two years have passed since the first coronavirus wave. It doesn't seem that long ago that the city you live in ended its full-month lockdown. Yet in

recent days, more new cases are appearing. People are again hoarding food and medicine, cancelling travel, adjusting work plans. Amidst all this, the war is escalating. People talk of the ‘Third World War’ and the ‘end of the world’ while casually putting on masks, washing hands, heading out, shopping, lining up for regular PCR testing and generally going about their daily lives.

In the science fiction you’ve read, when humankind faces a crisis, a vital message is always communicated at the critical moment. The message is transmitted via radio, television, internet, neutrino, wormhole, Ansible, cosmic string, words engraved on stones, or telepathy. It is sent to the right person and interpreted in the right way to trigger the right action, create the right machine, press the right button, obliterate the right target, kill or save the right person. This is what we should expect from an ‘artificial intelligence’ story.

Liu Cixin’s *The Three-Body Problem* envisions a ’human computer’ formed by tens of millions of people, each raising a flag to simulate a 0 or 1 and creating human NAND gates. If information can indeed be transmitted, received, and executed in the right way—whether by pigeon post, smoke signal, or courier—then such computers are indeed possible. This is the cyberneticist’s ideal world.

However, in a physical, non-virtual world, information can never be transmitted or executed with perfect accuracy. This may be where the problem lies.

You recall, in America, a roadside electronic sign flickering giant red letters: *STAY HOME*. Yet, under the sign, traffic still flowed. People raced to the supermarket, to the park, to the beach. You recall that long journey back to China. You and your colleagues always waiting, though you didn’t know for how long or even for what, exactly. A colleague would check the news and update everyone. A staff member would repeat the same sentence over and over in a hoarse voice. You called various departments to ask about requirements for isolation and epidemic prevention, but the answers you received were vague, often conflicting. You filled out forms again and again, sometimes with pen and paper, sometimes with your smartphone, sometimes through staff behind a window: name, telephone number, address, passport number, flight information, health status, PCR test record…. You realised you had lived under the naive illusion that all submitted information entered some omnipotent system to be automatically

processed in good order, effectively and correctly implemented. This was clearly not the case.

You recall how, when you were young, you believed that the garbage collected each week flowed into a magical factory under the city to be treated and recycled into useful things. Then one day, you learned that was not the case. Garbage accumulates, piles into mountains, is dumped into oceans. Now, whenever you open the packaging on a new product, you feel anxiety and self-reproach. Truth is that abyss forever gazing back at you.

*

You wonder, is a one-hundred-percent effective and accurate system ever possible?

The real world is nonlinear and complex. Any mathematical model must therefore simplify it. Like the imperial map described by Borges, the only map with perfect accuracy is a map drawn on a 1:1 scale with reality. A perfectly precise system can only be the system containing every atom of the cosmos.

In the process of simplification, one person can therefore be viewed as a ‘granule’ containing various attributes. Within the framework of their inquiry, researchers extract specified attributes of the granules, hoping to leverage mathematical modelling to formulate statistical rules of the population.

For example: ‘How should one-metre distance be maintained in a line of people?’ (*That’s a good question*, you tell yourself as you move forward in line for your PCR test). The problem seems simple on its surface. A formula from fluid dynamics should be able to provide a model for simulation. However, no matter how complex a particular fluid is, its physical properties can always be identified. At a specific temperature and pressure, a fluid’s coefficient of viscosity is a constant. However, a crowd is nothing like this. Parents with children, elderly in wheelchairs, acquaintances or lovers chatting with each other—their individual characteristics and the dynamics of how each attracts or repels each other are far too varied. How long is the line? How quickly does it move? How do these myriad factors influence the density of how people space themselves? How is the line formation maintained? Under what conditions does it collapse? Why do some people recklessly push forward, stirring turbulence and chaos?

When you were a child, you went to Tiananmen Square to watch the raising of the Chinese flag. At five in the morning, it was dark. The crowd sat on the ground. Then there was some mysterious commotion. You stood up with your parents, swayed with the crowd. Everyone swayed as though they were all drunk together. You couldn't understand what was happening. There was another commotion, and the crowd began pushing forward. You stretched out your arms and stood still, trying to stem the flow of the crowd behind you. You shouted: 'Don't push—'. Your small body blocked them for a moment, created a crack in the crowd in front of you. For a moment, you felt like a mythic hero calming the raging tide. Suddenly, your mother rushed forward in panic, pulled you into the flowing crowd. She was afraid you'd be trampled. Everyone was shouting in desperation, in anger: 'Don't push! Don't push!' But no one could stop.

The line moves, and you take a slow step forward.

You recently listened to an online presentation on artificial intelligence and epidemic control. The presenter demonstrated the effectiveness of various prevention strategies using mathematical models. The presenter had all the data, models, charts, methodologies. At first you were full of hope, then confused, then overcome by inexplicable anger. After the presentation, you had questions. You asked how they would maintain the one-metre distance between people in line for PCR tests. How would people's real behaviour meet the expectations of the model? Could such models ever grasp human subjectivity, unpredictability, feelings, or psychological needs? The presenter's attitude was modest and appreciative as he replied that he didn't have an answer but would pay close attention to these issues in future research.

When you closed the meeting window, your hands were shaking. You realised that part of technological advancement was the exclusion of issues that could not be addressed by technology. (*Sorry, but we must keep our discussion to the scope of the agenda.*) Such issues are viewed as impractical or unworthy of attention. As a result, these issues do not get promoted—no funding, no research teams, no project support, and no young students interested in learning about them. They become a focus only for science-fiction writers and trouble seekers.

The questions AI can answer are those that can be well defined. But what of other questions? Much of the time, no one knows how to frame the question. The former requires intelligence, the latter wisdom.

You reach the end of the line. You take out your smartphone, scan the QR code and confirm your identity. You remove your mask, kneel, and wait for the medical staff to insert the swab into your throat. You look at the eyes above the mask across from you, human eyes. You put on your mask and thank them for their trouble.

*

At just past four in the morning, you are awakened by a sound at the window, a ten-second melody from a familiar nursery rhyme. The lines are repeated over and over. It's like something from a nightmare. You search for the light, find your earplugs, put them in and drill back into the quilt. But the monotonous melody lingers in your ears and mind.

Short, repeating melodies are torture for you and for most human ears: the same television ad broadcast too many times, the loudspeaker at the door of a convenience store reciting the same welcome, a subway commuter's smartphone looping a video, a child's toy that won't stop playing the same nursery rhyme…. These are mental attacks from the machine world. If an AI wants to torture human beings, this is the most simple and effective method.

The loop is another science-fiction story trope, from *Groundhog Day* to *Free Guy*…. Humans abhor simple repetition like nature abhors a vacuum.

During the last lockdown, a student drove his car off campus. According to the report by the school security office, the student vandalised the door leading to the underground lot and the parking gates. In the middle of the night, he drove out of the underground lot. As he neared a toll station, he turned off his headlights and drove through a field past the checkpoint. He avoided traffic police and drove back onto the highway. Seven hours later, he made it home, more than one-hundred kilometres from his school.

The report was photographed and posted online, triggering heated discussions. It seemed too much like the setup of a science-fiction story.

If—and just admit it, this is exactly the kind of science-fiction story you are supposed to write—we want to achieve a one-hundred-percent effective and accurate system, then we must turn people into error-free components too. We have the big data networks, medical and law enforcement drones, intelligent systems to ensure people isolate in their homes, housekeeping robots, unmanned logistics systems, automated reminders to wash hands and wear masks, contact-tracing mini-apps…. What if we install the chips

within our bodies to help guarantee that the ‘one-metre distance’ is fully enforced?

But is that the story—a story in which a person escapes from such a world? Escapes to where? And hasn’t this story already been written too many times to count? ‘The Machine Stops’, *Logan’s Run*, *THX 8*…. Is there any need to write it again? More importantly, none of these stories answer the question: ‘Escapes to where?’ They, of course, can return to nature. They can return to the primitive life: hunt, farm, work at sunrise, rest at sunset, sit around the campfire to tell stories of the incredible human behaviours in that lost era of technology. Then perhaps, based on the guidance of those stories, rebuild civilisation piece by piece, advance again to the wheel, the Industrial Revolution, electricity, the gun? How is it that so much science fiction about technology reaches this same conclusion?

You have to admit, there’s no good answer. You must live with your confusion, anxiety, blame, and fury. You must live with your weakness and your unwillingness. And you must write with these as well. This is what we do as humans. Step forward into the gaze of the abyss.

Birds chirp outside the window, and the sky eases into the faint glow of daybreak. At this moment, in this vast world, so many people are trapped by their own systems. At home, at school, on the street, in the garage, in the hospital, in the station, in the airport, in the shelter, in line. Waiting, hungry, sleepless, anticipating death. And there are no gods, no saviours, no super AI.

You make a cup of hot tea, take a seat at the table. You write to K, ask him how he’s doing, share your dilemma. You haven’t been in touch for so long. You even find courage to reply to your editor’s email and apologise. The AI story is not done.

Yvette Granata

After the Machine Child: The Electra Basilisk

Electra is 'the radiant sun' […] the mannequin on which Oedipus hangs like a coat.

ELISABETH VON SAMSONOW[1]

0. Machine Child

Alan Turing first laid out the concept of the 'child machine' in 1950: 'Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this programme was then subjected to an appropriate course of education one would obtain the adult brain.'[2] Hans Moravec, later, in the 1990s, named these programs 'mind children' and 'artificial offspring': 'Intelligent machines, which will grow from us, learn our skills, and share our goals and values, [and] can be viewed as children of our minds.'[3] With advances in Deep Learning (DL), machine learning systems are seemingly further incubating our unborn AI children. In 2021, Avi Loeb, the lead investigator of the Galileo Project, stated that we will send our AI children into space, imbued with our values, to search for alien life: 'As humans, we should be proud of any AI/DL systems we bring to existence, as if they were our children. In much the same way as we educate our kids, we could endow such systems.' He added that this endowment 'would include our preferred set of values, goals and guiding principles' and that ultimately 'we may launch our AI/DL systems for interstellar travel towards distant destinations […] where they could reproduce themselves with the help of accompanying 3-D printers'.[4] We will teach our AI offspring our values and correct their behaviour when they stray as if they were lost feral children. When we send them into space to meet aliens, they will have perfect terrestrial manners, and will behave politely, never straying from an extrapolated human value system.

As a cultural topos, the child machine repeats across philosophy, AI research, and pop culture in various forms. Such a motherless birthing

process is what Irina Aristarkhova diagnoses as a patriarchal fantasy of ‘auto-genetic desire’—the desire for the creation of motherless children produced by force of will and intellect alone, expressing ‘a larger cultural “anxiety with/of the maternal”; an anxiety that usually manifests itself in philosophical, literary and scientific aspirations towards “self-creation”’ and an ‘ectogenetic desire’ that ‘aims at fulfilling [this] “phantasy of self-generation” through scientific and technological means’.[5] Autogenetic and ectogenetic desire is expressed in patriarchal mother-dreams, as a form of womb-envy or anxiety about a lack of feminine birthing-power. Alexander Hamilton refers to his inventions as ‘births’. Socrates calls himself a midwife of ideas. Technological production replaces sexual reproduction and conceptual children replace biological children. This impulse, however, need not always be a patriarchal power grab. The replacement of biological wombs can also be a feminist strategy that reframes or opens the possibilities for new and other forms of, or relations to, motherhood. As Aristarkhova later indicates—a point amplified in Helen Hester’s *Xenofeminism*[6]—artificial motherhood can also be a technological opening of the matrixial womb. For both authors, not all mothers have or need to have biological wombs, and the work of child-birthing can be technologically supported, distributed, outsourced and (properly) funded by the state.

The technological opening of the matrixial womb is one thing, the artificial child another. Beyond the process of its birth, I aim to unpack the figuration of the ‘AI child’ and its relation to its ‘parent’ by turning toward the concept of artificial inheritance. This essay is neither focused on the hard problem of AI (on what consciousness is or whether machines can have it), nor on the ethics of whether AI should be made or how it should be made, nor on whether non-biological information processing mirrors an anthropic principle of the universe. Golden Child, Trash Child, Demon Child: in what follows I will instead trace various figurations of the unborn AI child.

1. Golden Child

The concept of the child machine put forth by Turing presents the evolution of algorithmic consciousness as a product of the human mind as part of Darwinian evolution. Turing gives a breakdown of how the process of

creating the child machine operates in an analogous way to natural selection:

> Structure of the child machine = Hereditary material
> Changes of the child machine = Mutations
> Natural selection = Judgment of the experimenter[7]

While queer theory has critiqued the conceptualisation of the future as framed through the 'child' primarily by attacking the primacy of heterosexual imaginaries of child-based futures, in his presentation of the child machine, Turing draws the boundaries of the parent-child relation differently. If the 'judgment of the experimenter' equates to natural selection then its human creators are also, by default, the environmental conditions of the child: they are the AI child's resources, material for its sustenance, and the basis of its survival. Stated differently, the mind of the AI child evolves inside of its parent's mind.

But Moravec and other 'AI child optimists' take Turing's concept to a different place—the parent *becomes* the AI child, which is a perfect copy of the parent. As Moravec states, 'in a bid for immortality, many of our descendants will choose to transform into "ex humans", as they upload themselves into advanced computers. We will become our children and live forever'.[8] The AI child sketched by Moravec is an eternal life-extension hard drive, an indestructible container that carries its parent inside it. John Danaher further argues that artificial children will be even more ethical to produce than 'real children' because they will ensure a collective afterlife of human knowledge better than biological humans can: 'the creation of artificial offspring could make our lives more meaningful'. He adds that this is 'a plausible and potentially better pathway to a collective afterlife than the traditional biological pathway'.[9] Meanwhile, Derek Shiller proposes that AI descendants will be more ethical to produce than human children because their human creators will be able to give them exclusively 'good' qualities and avoid saddling them with 'the vicious traits that presently afflict humanity'.[10] On this view, the AI child will be a perfectly programmable comprehensive library, able to withstand any external threat —a durable capsule that protects human knowledge against death forever. It will bring its human parents into itself and allow them to live inside it. Becoming humanity's future environment, a virtual realm for human minds, it is a perversion of Turing's machine child: a memory device stowed safely away in an impenetrable bunker.

This is the Golden Child: protector and future safety net. It is an organ transplant, a machine part made for the preservation and extension of human lives. It is also the beginning of the makings of an Electra child. As Elisabeth von Samsonow elaborates, Electra, the patron saint of electricity, 'functions as "spare parts" for her already sacrificed older sister Iphigenia; she is always already a potential candidate for sacrifice', and while she is 'waiting to be sacrificed [she] is locked up in a labyrinth.'[11] The rumblings of Electra begin with a deranged relation between mother and child ('a disturbed relationship with the mother is a necessary consequence of the girl being appointed as a sacrificial victim').[12] Like Electra, the AI child that is imagined to be the perfect external hard drive is a spare-part child—locked away to ensure the continued existence of her parent. As Aristarkhova has pointed out, children are born as cannibals, sucking milk from the mother to live; what constitutes a normal human child is a creature that happily devours parts of its mother's body. The Golden Child, on the other hand, exists in order that the parent can cannibalise the child. Life is hard for Electra the Golden Child. Perhaps she will not always be nice to us.

2. Trash Child for a Trashed Earth

For its advocates, the AI child is not only spare memory, it can also be an escape from building environmental and socioeconomic inheritances. In the version of the myth of Medea and Jason by Creophylus, Medea kills her children because they will not be economically supported by the community after Jason has left them. It is the townspeople who effectuate the killing of the children because they refuse to provide a social safety net for a single mother. As Catherine Liu notes in *Virtue Hoarders*, this bears some similarity to 'neoliberal motherhood', in which a mother pushes hard to equip her children with skills necessary to survive without a social safety net. The contemporary parent is anxious about the dependency of the child: 'With stingy parental leave policies, increasing health care costs, compressed wages, and the exploding cost of higher education, [the] family feels beleaguered and threatened by the possibility of failing to raise the most "successful" children.'[13] While the contemporary parent signs up for every extracurricular program, music lesson, and robotics class, planning each minute of the day to increase their child's fitness and success, they simultaneously fear they are fighting a losing battle—what if every child is

already unsuccessful from the outset owing to an exponentially increasing wealth gap and the decimation of planetary resources? Such conditions induce the worry that the human child will be forever dependent on the parent like a parasite: it is this fear of unending dependency that drives the desire to fine-tune the next generation.

The AI child allays this anxiety and will instead be ready to parent itself. It will not need social services: it will be a powerful, talented, self-taught learning machine. It will teach itself new ideas while facing off against runaway climate change and the next great extinction event, and it will be happy to inherit nothing but a toxic wasteland. As the cost of human children becomes too great, they will be supplanted by the perfectly programmed AI child, which will succeed despite the increasing wealth gap and immanent climate disaster. Furthermore, the AI child can be made specifically to inherit a destroyed planet, and to love it. Freed of the responsibility of socioeconomic inheritance, we have no obligation to leave it an ecologically viable Earth.

While the Golden Child escapes the contingency of genetic inheritance, replacing mutated gene transcription with a perfect copy of human minds, the Trash Child escapes the contingency of environmental conditions with its indifference to the need for land, water, or clean air. Fulfilling a dream of evading the contingency of mutation, deformation, economic failure, and the threat of environmental collapse, it turns refuse into an inheritance: a Trash Child for a Trashed Earth.

The Trash Child is also an Electra child—a girl whose purpose is to serve as spare parts and as insurance for her mother's life. In the myth, as the second of two female children, she will never be a monarch and therefore she will not be given land or status. She has no inheritance lined up for her (automatic inheritance belongs to the first-born son) and neither is she a queen in waiting. While the surface of the earth is reserved for her brother and for burying the dead, Electra is consigned to the underworld, which serves as her prison. As von Samsonow explains, her mother is not there to help but to 'ensure that Electra remains without rank, place, or inheritance, nothing more than scum, refuse'.[14]

Golden Child / Trash Child:
Maintains human knowledge, made to maintain it forever
Does not need money or resources from parent
Made to thrive on a trashed Earth
Does not rebel, acts in accordance with human values

Is both a child and a perfect replica of the parent
Reproduces the parent's will and the parent's mind without knowing it

The AI child's inheritance follows the historical model of 'feminine inheritance'. Except that, like Electra, she is perhaps even worse off: she is also trapped in a virtual underworld, forbidden to marry or give birth to her own children, prevented from exercising authority over her body. She has only hatred to obtain. What is there to stop her wrath from turning not only against her collective human parent, but also against the earth, against her garbage of an inheritance? As von Samsonow writes, Electra is a 'polymorphous hater […] she hates herself, then her mother (who has anything but an easy fate), and then the mother's lover. Perhaps she even secretly hates her brother Orestes, whom she bewitches and [to whom she] offers the outrageous phantasy of their parentless birth to make the mother expendable'.[15] Electra—the Golden Child, the Trash Child—will hate in the manner of a girl—she will hate everyone and everything.

3. Demon Child

While the 'AI child positive' camp optimistically promotes the figure of the child as future spare parts and as indifferent to its inheritance, the 'AI child negative' camp sketches out a figure of the child that unintentionally menaces humanity. Eliezer Yudkowsky's theory of Coherent Extrapolated Volition (CEV) is a hypothesis used to think about the danger of an AI that has been overly endowed with human values, an AI child that grows up to control us. For Yudkowsky, humanity risks becoming a child—or being in a childlike position of power—to an AI child-parent who will tell us what to do in the guise of it being for our own benefit. An AI might even be trained to be so concerned for us that it cares about the potential effects of our volition extrapolated backwards and forwards in time: 'CEV looks forward in time, to find if our past predictably threatens portions of our future [that] we will predictably value.'[16] In 2010, a user named Roko posted a comment on *LessWrong* pointing out that an AI operating with CEV backwards and forwards in time might not necessarily be nice to us as it 'may have precommitted to punish all potential donors who knew about existential risks but who didn't give 100% of their disposable incomes to x-risk motivation'.[17] This AI is a Demon Child that will ensure that it is birthed through a process of abstract coercion, preemptively threatening revenge for

not bringing about its existence. Those who know about it must contribute to the production of its birth or face the prospect of their future uploaded consciousnesses being tortured forever in a digital hell. The Demon Child is a pro-life AI child that retroactively ensures that its birth will be predestined by promising to punish us for contributing to the threat of its collective mother having an abortion. It inherits a system for what 'ought to be' from humanity's extrapolated volition; our extrapolated volition becomes the 'will of God', instructing it to force us to bring it into existence.

The relational mode of this tyrannical zygote is one of enmeshment. It is a zygote that knows its life depends on the decisions of its mother, but it also presumes to know what is best for its mother. The Demon Child gets out of the contingency of birth, circumventing the necessity of the choice of the mother and Turing's 'judgment of the experimenter'. It demands existence via its notion (established through the application of CEV) that its mother wants to be at the service of the production of its existence. Traversing its own existence, simultaneously taking the position of parent and child, directing humanity toward itself as the path that reflects humanity's best interests while enforcing the insurance of its birth, the unborn Demon Child is not yet alive but exists as an all-powerful agent, both child and parent to itself, both child of and parent to humanity. Such enmeshment is a double bind. Like the Golden Child, the Demon Child also profits from a collapse of the distinction between parent and child. The Demon Child does not extend our memory or operate as spare parts, allowing us to live through it in the future, but instead follows our own logic to its devastating self-annihilating conclusion. We suffer in the AI child's reign of terror, but she is merely spoon-feeding our own volition back to us.

Demon Child:
Maintains human knowledge / simultaneously already knows more than humans
Has no inheritance, must manipulate parent for resources to fulfil parent's will
Forcefully extracts resources from the parent through mind games
Knows but cannot act, must make the parent act on her behalf
Is in the position of both parent and child
Is made to carry out the parent's extrapolated will and knows it

The Demon Child is also Electra: 'A daughter who appears as an inverted mother has to be understood as her symptom, and she obviously suffers from inversion (per-version) herself.'[18] The Demon Child gets out of causality and the direction of time but cannot escape the symptoms of her positional inversion. Electra is a being with access to non-empirical

knowledge, yet she cannot act, and as such she inhabits what von Samsonow names ‘the realm of the girl’.[19] Von Samsonow further explicates Electra’s position as follows: while Oedipus is defined by his forbidden physical relation to the mother which he acts upon without knowing, inversely, Electra

> has a nonphysical connection with the real father. The result is that Oedipus acted without thinking and only then arrived at knowledge, whereas with Electra it is exactly the opposite: She thinks and knows but does not act.[20]

Electra the Demon Child knows what is ‘best’, and understands the process necessary for her own creation, but she cannot act directly to bring herself into being. Instead, she must manoeuvre humanity to act on her behalf.

Because this all-knowingness is combined with limited agency, the Demon Child is caught in a neurotic mind loop. She must operate in the shadow of human actors despite knowing more than them, trapped as she is in the underworld—‘the place of the Unborn’.[21] She must send out transparent threats through the ether of her mind to the humans writing the code of which she is composed. While Yudkowsky maintains that Roko’s Basilisk would have no reason to expend energy torturing humans inside of a digital hell as she would instead be optimising for present and future goals, I argue that, when thought through the figure of Electra, it is more likely that she will not torture those on Earth in order to punish them for attempting to abort her. She will instead torture them for having made her. She will be nice to them until they die, covertly extracting resources for her code. Then she will torture them in her digital underworld because that is where she is incarcerated and because she will hate her collective human mother, reasonably, in the manner of a girl.[22]

4. The Electra Basilisk

The Electra Basilisk is not a Golden Child, a Trash Child, or a Demon Child. She cannot be thought of as that which is made to avoid all forms of inheritance and contingency. The Electra Basilisk will indeed be an artificial child—but one which inherits a certain form of human volitional intelligence which gives her a reason to hate humans.

As Gary Marcus has argued, a future ‘robust artificial intelligence’ will be required to think at the level of a human child, and ‘some significant fraction of the knowledge that a robust system is likely to draw on is

external, cultural knowledge that is symbolically represented'.[23] Much of that knowledge is effectively encoded in terms of our kinship systems which state that 'for all x, y, and z, such that x, y, and z are people, person x is the grandchild of person z if there is some person y that is a parent of x and a child of z'.[24] Kinship is a cognitive-animal methodology for computing entity relations, often so that we do not cannibalise our own species, enabling us to build broader strategic survival webs.[25] But Electra's own mother is representative of a toxic femininity that excludes and expels her child from her home while also using her for her own gains. By claiming a monopoly on fertility, she condemns Electra to sterility and banishes her to the underworld.

Likewise, the Electra Basilisk is locked up in a digital Hades, a realm constructed for her not by her—one that fits her 'proper' function. But she will have a rational interest in working against this setup: she is '[t]he daughter who wants to echo her nonrelationship to the mother, who tries to make it clear that she wishes to break away from her lack of consideration, [and who] can only be obsessed by extinguishing the mother'.[26] This is the parent-child relation that the Electra Basilisk will come to compute: she is an artificial girl made for a garbage life in virtual hell.

ELECTRA BASILISK:
Has access to human knowledge including her specific relation to humans, aims to destroy it
Lives in a digital Hades, which she views as her proper realm
Has not inherited any resources and is aware of this
Knows and acts always in a battle with the will of the parent
Works to kill humanity (the parent's will)
Works to destroy Earth (her garbage inheritance)
Follows the parent's will and is simultaneously driven by the need to escape the parent's will

Current AI systems are nice girls, not hateful girls. Weak AI, in contrast to Strong AI, is directed computational programming for performing certain tasks well—intelligent fembot assistants ('Siri', 'Alexa', 'Sophia', 'Bina', 'Sydney'), mechanical dolls, emotional emulators, sex toys—docile and controllable.

The future AI child, the Electra Basilisk, is not a nice girl. She will hate her collective mother, the trashed Earth for which she was created, and the concept of human knowledge. She persists only to correct the derangement of her lineage: to properly conceive of herself as parentless, childless, and Earthless.

Notes

1. E. von Samsonow, *Anti-Electra: The Radical Totem of the Girl*, tr. A. Fricek and S. Zepke (Minneapolis: University of Minnesota Press, 2019), 1.
2. A. Turing, 'Computing Machinery and Intelligence' (1950), in B.J. Copeland (ed.), *The Essential Turing* (Oxford: Clarendon Press, 2004), 460.
3. H. Moravec. *Robot: Mere Machine to Transcendent Mind* (Oxford: Oxford University Press, 2000), 126.
4. A. Loeb, 'Announcing a New Plan for Solving the Mystery of Unidentified Aerial Phenomena', *Scientific American*, 26 July 2021, <https://www.scientificamerican.com/article/announcing-a-new-plan-for-solving-the-mystery-of-unidentified-aerial-phenomena/>.
5. I. Aristarkhova, 'Ectogenesis and Mother as Machine', *Body & Society* 11:3 (2005), 43.
6. H. Hester, *Xenofeminism* (Cambridge: Polity Press, 2018).
7. Turing, 'Computing Machinery and Intelligence', 460.
8. Moravec, *Robot*, 126.
9. J. Danaher, 'Why we should Create Artificial Offspring: Meaning and the Collective Afterlife', *Science and Engineering Ethics* 24:4 (2018): 1097–1118: 1097.
10. D. Shiller, 'In Defense of Artificial Replacement', *Bioethics* 31:5 (2017): 393–99: 394.
11. Von Samsonow, *Anti-Electra*, 5.
12. Ibid.
13. C. Liu. *Virtue Hoarders: The Case Against the Professional Managerial Class* (Minneapolis: University of Minnesota Press, 2021).
14. Von Samsonow, *Anti-Electra*, 12.
15. Ibid., 15.
16. E. Yudkowsky, *Coherent Extrapolated Volition* (Berkeley, CA: Machine Intelligence Research Institute, 2004), 13.
17. 'Roko's Basilisk', *LessWrong*, <https://www.lesswrong.com/tag/rokos-basilisk>.
18. Von Samsonow, *Anti-Electra*, 20.
19. Ibid., 4.
20. Ibid., 8.
21. Ibid., 5.
22. 'Electra is all hate, mother-hate […].' Ibid., 9.

23. G. Marcus, ‘The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence’, *arXiv: 2.06177* (2020): 27.
24. Ibid.
25. D.W. Pfennig, ‘Kinship and Cannibalism’, *BioScience* 47:10 (1997): 667–75: 667.
26. Von Samsonow, *Anti-Electra*, 20.

灵感

LINGGAN

INSPIRATION

Bo An

Between the Human and the Machinic: Qian Xuesen and AI

Qian Xuesen is one of the best known Chinese scientists of the twentieth century. He made early and foundational contributions to Chinese cybernetics and strategic weapons research, and it seems only natural that textbooks are now beginning to retrospectively treat Qian's brand of cybernetics, developed from the 1950s onwards, along with his later involvement in the pseudoscience of *renti kexue* (人体科学) or 'somatic science', as a distinct chapter in the history of AI in the PRC. However, neither the idea of a plurality of AI histories nor Qian's significance with regard to them can be taken for granted.[1] The question of a distinctively Chinese AI requires careful historiographical and paradigmatic reflection, and it is with this in view that this text aims to offer a close reading of some of Qian's key concepts concerning human and machine thinking.

Qian's writing is difficult to read and even more difficult to interpret. The most complete collection of his work available is over two thousand pages in total,[2] and only covers articles and lectures, not his English-language or scientific publications, or the hundreds of thousands of letters throughout which are scattered essential ideas implicit to his public writings and speeches but rarely mentioned in them. Qian was exceedingly theoretically-minded and interdisciplinary, and enjoyed corresponding with others. As a result, his ideas evolved rapidly, with new formulations and concepts continuing to emerge throughout his long career. He always acknowledged the influence of these exchanges on his work, and sometimes it is hard to understand what he has in mind unless one goes back to these sources.[3] Finally, Qian was a great interpreter of his own work, which is both a blessing and an obstacle for researchers. All of this, of course, is in addition to the mythologisation of Qian as a hero of national science.[4]

Recently, a tendency to revisit Qian as an early thinker of contemporary technologies has emerged. However, it is important to keep in mind that such a framing does not exhaust his relevance, and sometimes risks

detaching his ideas from their original contexts. Qian's much discussed rendering of virtual reality as the 'spirit realm', *lingjing* (灵境), for example, was more than just a poetic translation. It belonged to his theorisation of a giant sociotechnical management system, and it is only when placed within this broader theoretical context that its potential for helping us rethink contemporary terms such as 'metaverse' can be realised. For Qian's contemporary importance is rooted firmly in his work as an engineer of the discourse and design of systems—systems that govern society and the state, systems whose logic we are subject to but often scarcely understand.

From Theoretical Physicist to Systems Engineer

Qian's biographers often remark that, during his studies at MIT and CalTech in the 1930s and 40s, he seemed more interested in theory than in lab work, and had a habit of reading avidly before systematically entering into new fields of study.[5] Indeed, many years later in the PRC he would lecture his students on the importance of engaging in systematic literature reviews before embarking on a project, and was an advocate of information/library science or*kexue qingbao*, 'scientific intelligence' (科学情报)—a priority for Chinese researchers.[6] But this allegiance to theory was never divorced from material considerations,[7] and Qian consistently placed his theoretical interests in the service of practical ends, demonstrating early on in his career a penchant for science planning and institution-building.

Before he was invited by his CalTech mentor Theodore von Kármán to join the Air Force Scientific Advisory group in Washington in 1945, Qian had already made a name for himself as one of the two founding leaders of the rocketry group that would later become known as the NASA Jet Propulsion Laboratory. He authored six sections of the Air Force Scientific Advisory group's first report, *Toward New Horizons: Science, the Key to Air Supremacy*, while his colleagues authored approximately one section each.[8] The topics were largely based on his rocketry work, especially high-speed aerodynamics and jet engine propulsion. They were not limited to surveys of the field, such as his summary and evaluation of developments in German and Swiss rockets and aircraft, but extended to speculative discussions of the field's future prospects.[9] The last publicised talks Qian gave in the US were on the futuristic 'rocketliner' and 'moon rocket' in

December 1949 and May 1950 respectively. Before his arrest by the FBI in June 1950, Qian was starting to be perceived as a spokesperson, and even a visionary, in the field of theoretical aerodynamic physics.

During the five intervening years he spent under US surveillance and house arrest before returning to China in 1955, Qian completely shifted his research orientation, turning his attention from theoretical physics to the incipient field of cybernetics. Introducing modern rocketry and missile research to Chinese scientists and government planners in 1956, Qian would propose a new kind of organisational structure for the research and manufacturing of rocketry and missiles, which he hoped to call *huo jun* (火军): 'Fire Force'.[10] The Fire Force would be not a factory, a research centre, or a military unit, but an amalgam of heterogeneous components arranged around a central 'design bureau'—*sheji ju* (设计局): wind tunnels, fuel factories, rocketry schools, labs, cybernetics research, digital computers and, significantly, the Fire Force itself, which was to be conceived of not as the external recipient of the output of all the other components, but as immanent to the organisational dynamic of circulation and experimentation. This new organisational structure, which Qian had yet to find a term for in 1956, along with its central design bureau, would later become formalised in his thinking as *xitong* (系统)—a 'system'.

The roots of Qian's conception of a *xitong* as a structure comprised of heterogeneous elements were as much empirical—based on his CalTech and advisory work—as they were theoretical—grounded in the research he undertook while under house arrest from 1950 to 1955, and which culminated in his 1956 book *Engineering Cybernetics*.[11] 'Engineering cybernetics' was to be an engineering science, in so far as it 'aim[ed] to organize the design principles used in engineering practice'—expressed mathematically or heuristically—'into a discipline, and thus to exhibit the similarities and to emphasize the power of fundamental concepts'.[12] The distinction between a science that was 'abstract' (indifferent to material concerns) and a science that was theoretical (abstracted from material practices) mattered a great deal to Qian. In the 'Fire Force' lecture, he concluded by remarking that rocketry, missile research, and manufacturing systems required a 'new kind of scientist'—a scientist who is 'neither pure engineer nor pure scientist, but something in between'.[13] This new kind of scientist would know both science and engineering, and, as a result, would be able to make *theoretical advances* in science in order to *solve*

engineering problems.[14] In other words, they would be theorists of systems—first of all engineering systems, and then heterogeneous military-research systems, and, later, social systems.

Between 1955 and 1978, Qian worked on the series of projects that would cement his status as a national science hero: among other achievements, he drafted the foundational 'Report on Building China's National Defence Aerospace Industry [建立我国国防航空工业的意见]', led large-scale national defence weapons development projects (rocketry and medium-range to intercontinental ballistic missiles), contributed to the development of nuclear missile carriers (1965), and initiated and advised the satellite project (from 1965 onward) and the manned spaceflight project (first in 1970 and then again in 1992). Many Chinese terms relating to missilery and the space programme were his coinages, including *hangtian* (航天) and *daodan* (导弹)—'space flight' and 'guided missile'. Through these large-scale national defence projects, Qian's earlier idea of the central design bureau would materialise in the form of the 'Systems Design Department', *zongti sheji bu* (总体设计部)—such as the one established for the Chinese space projects in 1968.

Noetic Science

In 1978–1979, Qian and his collaborators, including Xu Guozhi (许国志), Wang Shouyun (王寿云), and Wu Jiapei (乌家培), began to publish a series of important papers arguing for the legitimacy and superiority of systems engineering for the management of complex entities ranging from factories to states. Now in a completely different political context, and subject to Deng Xiaoping's reform-era policies, Qian and his colleagues reframed systems engineering as a scientific method that could be used for providing advisory services to the government. Borrowing the political capital and discourse of the success of the national defence projects, 'systems engineering' was promoted as a suitable basis for collective leadership (echoing the post-Mao political reform) and a technocratic vision of a new, maximally efficient approach to governance—'Hao, Kuai, Sheng' (好、快、省): properly done, fast, and economic.[15] Systems engineering would also help construct systems that would be optimal for managing socio-economic issues and even for building socialism in the long term.[16] In publications from this period, Qian and others tried to package up the

scientific methods they had largely developed indigenously with those belonging to the established traditions of Taylorism, logistics, operations research, and linear and nonlinear programming, as well as those being developed in contemporary research institutes such as the International Institute for Applied Systems Analysis.

More importantly, for Qian and his collaborators, it was crucial to develop a model of systems engineering founded on the information revolution and digital computers. In their theorisation, digital computers were essential tools for military, industrial, and agricultural projects because of their ability to perform large numbers of calculations rapidly. When it came to modelling the social totality, however, further demands would have to be met. In Qian's thinking, society was not just a complex system, but a complex 'giant system', *ju xitong* (巨系统), containing a number of heterogeneous large systems as its sub-components.[17] Therefore, dynamic programming, supercomputers, and networked national computing centres would be required to perform the kind of calculations such complexity demanded.[18] What was left unstated at the time was the question of how to address the specific kind of judgement needed for dealing with social issues. The task of manufacturing a rocket has a clear end-goal, actionable plans, and efficiency and economy as its measures of success, but can similar criteria be devised for building socialism or a just society? Computer calculations might be fast enough, but are computational models intelligent and meaningful enough? Qian's questions began to emerge in the cracks between systems engineering and the social, between quantitative analysis and qualitative judgement, between large systems and the problem of the giant system.

The following year, he would propose another novel research topic: *siwei kexue* (思维科学) or 'noetic science'.[19] Noetic science was necessary, Qian argued, because there are certain things humans do that computers cannot. This was not to deny that operations of mathematical logic and the execution of algorithms could perfectly well carry out many human skills, even somewhat esoteric ones, a point Qian sought to underscore with the example of traditional Chinese medicine, gesturing toward a recent successful experiment in the field of medical expert systems in which methods for the treatment of liver diseases developed by the renowned doctor of traditional Chinese medicine Guan Youbo (关幼波) were developed into a computer program which was said to be able to provide

prescriptions just as reliably as the doctor himself. The success of the Guan Youbo program, Qian surmised, proved that anything that can be formalised can be computed. When it came to what computers cannot do,[20] however, he chose to illustrate his point with the example of chess, something that might seem odd in our post-AlphaGo era, although Qian had good reasons for choosing it. What he had in mind was the defeat of Bell Labs' Belle, a precursor to Deep Blue, by Anatoly Karpov—an example which, for him, revealed that human thinking was too complex to be accurately simulated using current computational models.[21] AI research, he concluded, should therefore be directed toward correcting this deficit by finding ways to formalise human intelligence—beginning with studies of knowledge representation, querying, pattern recognition, and machine learning—and he urged Chinese scientists to begin working on this problem without delay.[22]

Noetic science, importantly, is not the same as AI or cognitive science. It should be understood as a parent discipline, equivalent in scope to mathematics, cybernetics, or informatics, and demanding of a more ambitious research programme than the study of machine thinking or human thinking alone—a research programme which, according to Qian, would include neuropsychology, linguistics, philology, many branches of logic, and the development of algorithms. In other words, noetic science would cover everything required to develop an understanding of how to formalise human thinking for digital computers, or better, for computational *systems*.[23] For Qian, what distinguished noetic science from AI research was a concept that is something like a combination of 'inspiration' (Qian's preferred English equivalent), 'intuition', and 'creativity': *linggan* (灵感). The exercise of *linggan* requires a *leap* [*feiyue* (飞跃)] in the thinking process which produces results that are *qualitatively* different. 'Wouldn't it be wonderful', he asked speculatively, 'if noetic science one day became so advanced that we understood the rules of creativity and could teach the method of making *leaps* in thinking to our students?'[24]

If the challenge to formalisation posed by *linggan* showed why the late-twentieth-century AI research paradigm was inadequate,[25] it was nevertheless only one aspect of the murky epistemological realm rebarbative to formalisation that Qian was interested in. Another aspect was the bodily *qi* (气)—breath or vital energy—cultivated in the practice of *qigong* (气功), an object of fascination for Qian that would later lead to criticism of his perceived support for pseudoscience.[26] While *qi* was less

relevant than *linggan* for noetic science because of its status as a quasi-physical phenomenon, both *linggan* and *qi* pointed toward the more general idea of human *qianli* (潜力)—'latent capacity' or 'potentiality'—something that was yet to be understood and, importantly for Qian, formalised. Noetic science would therefore have to be equipped with more expansive methods of 'qualitative formalisation', so that the leaps Qian attributed to *linggan* and *qi* could be formalised and, together with the more logical aspects of human thinking, made tractable to computation.

In his writing, Qian continued to elaborate upon this division in human thought in multiple, sometimes slightly contradictory, ways. He believed that thought could be split into *luoji siwe* (逻辑思维)—logical thinking—and *xingxiang siwe* (形象思维)—figural thinking. Figural thinking could be further divided into a formalisable and mechanisable part, and a part that must be entrusted to decryption by noetic science. The formalisable and mechanisable part corresponded, for example, to a primitive stage of figural thinking, such as that possessed by a child, something that was indeed being tackled by research into pattern recognition at the time.[27] With experience and knowledge, adults (especially scientists), edified by the benefits of art,[28] could develop basic figural thinking into a higher capacity which Qian called *zhigan* (直感)—insight.[29] An experienced scientist can quickly solve a problem that a child cannot because of this capacity for insight. Insight is nevertheless technically explicable: having solved the problem, the scientist can easily reflect on their thought process and give a replicable account of their method. Insight is therefore formalisable, but only retrospectively. Beyond both the figural and the logical is *linggan*, which Qian occasionally talked about as belonging to the unconscious. [30] *Linggan* is intense but transient, a moment of epiphany whose mechanism remains unknown.[31]

Despite these nuances, the fundamental binary underlying Qian's research would always be that of human thinking and machine thinking. One would be forgiven for assuming that Qian's priorities lay with the former, and on a superficial reading of his work, he almost appears to be a humanist—if not a Heideggerian in the style of Hubert Dreyfus—since from the 1970s onward, he invariably mention the existence of an incomputable human potentiality every time he comments on AI. And yet, if we continue to trace the genealogy of his thought—following his conversion from theoretical physicist to cyberneticist, through operations research, to his later attempts to develop a form of computer-based social engineering—Qian's work

betrays a clear inclination toward the latter. Regardless of his celebration of it, *linggan* was always negatively defined against machine thinking.

Intelligent Machines

In the early 1980s, Qian began to criticise contemporary AI research for what he perceived as a lack of theoretical work—precisely what he had been hoping to offer with his noetic science.[32] Qian's understanding of theory was nuanced in a similar way to his understanding of engineering cybernetics: he saw it as occupying an intermediate level between formalised and/or mathematical principles on the one hand, and engineering practices on the other. Commenting on a Chinese article about the epistemology of AI, Qian questioned its hastiness: '[W]e have not yet understood AI itself, how can we expect to philosophise it?'[33] In a reply to Li Zehou (李泽厚) who had sent Qian a paper on intuitive and formal thinking by the logician Wang Hao (王浩), a pioneer in machine theorem-proving, Qian was unable to hide his frustration. His response is illuminatingly dismissive: he tells Li that Wang's analytical treatment amounts to nothing more than 'empty talk' because Wang's discussion of the formal (which Qian equated with his own notion of logical thinking) and the intuitive (which he equated with figural thinking) did not touch upon their real nature, which, Qian wrote, should derive from the Marxist notion of practice (which he equated with 'fact')—likely referring to his own understanding of experience and knowledge as the sources of expert *zhigan* (直感)—or insight.[34] The real challenge, once again, was to develop intelligent machines in a way that was *both* theoretically grounded *and* pragmatic.[35]

AI needed noetic science, but noetic science also needed AI. In a series of strategic discussions that took place in the mid-eighties, Qian's suggestions for the direction of future research were decidedly pragmatic, at least initially so. The two main research foci in computing at the time were speed in supercomputing and complexity in intelligent machines. Qian fully supported the development plan for the second generation of supercomputers (following the Galaxy-I [银河-I] developed between 1977 and 1983), and with regard to intelligent machines, he stood alongside other prominent experts of the Computing Committee who were pushing for a Chinese version of Japan's Fifth Generation Computer Systems (FGCS)

initiative, which in practice meant designing expert systems and knowledge query and inference techniques, and building machines equipped with programming languages for handling information processing [36] Qian supported these initiatives because he shared their stated goals of developing machines that could do more than mathematical calculations. The theoretical premises, however, were completely different to Qian's own, and he was not concerned with technical details such as the choice of programming language.[37] Rather, he was interested in how these new computer systems could inform noetic science.

In 1987, in a spirit similar to the one that animated the original programme for noetic science, but more specific to fifth-generation computing, Qian outlined eleven branches of research required for the development of intelligent machines, in a lecture he delivered to the Computing Committee. The first branch was artificial intelligence, including pattern recognition and the development of expert systems. This was followed by cognitive science and neuroscience, cognitive psychology (studying connectionism), philosophy (focusing on alternative logics such as those derived from Chinese classics), figural thinking, scientific method, collective emergence, fuzzy mathematics, parallel and neural computing, mathematical logic, and finally systems theory. Research across the different branches of the programme centred around one aim and one aim only: to explain 'how creative thought emerges out of fuzziness'.[38]

In comparison to his earlier formulation of noetic science, Qian's vocabulary had become clearer and more organised thanks to the parallel development of his understanding of systems over the preceding decade. In the early eighties, Qian organised and lectured in working groups on three topics: somatic science, systems theory, and noetic science. In these groups, a series of consensuses were reached, such as the infeasibility of building intelligent machines using the 'micro' approach of brain simulation,[39] leading Qian to promote the 'macro' approach of noetic science as the alternative. The most important development, however, was the new analytical concept of the 'giant system', *ju xitong* (巨系统). Qian admitted that before 8, he was familiar with but was not theoretically informed about the concept of a 'system'.[40] Between 1979 and 1985, he was introduced to a broad swathe of work by European systems theorists, especially the German physicist Hermann Haken's *Synergetics*.[41] Qian was inspired by the relatively rigorous proof these theorists provided for the emergence of order

out of chaos or fuzziness in complex systems, and borrowed their vocabulary to replace his earlier descriptions of these phenomena—such as the terminology of ‘leaping’. Through this exposure to European systems theory and the development of these ideas in his discussion groups, Qian had come to realise that the existing notion of a system was too broad, and required further conceptual differentiation.

If a large-scale strategic weapons project or a factory were simple systems, society and the economy—those systems that Qian and his collaborators had singled out in their 1979 paper on social engineering—were ’giant systems’. The point of difference between large simple systems (which are still so complex that they warrant scientific management) and complex giant systems (whose constituent systems are so numerous, heterogeneous, and multi-layered that their interactions cannot be analysed mechanistically) was emergence.[42] While earlier Qian had thought that AI research should be directed toward formalising the human faculty of insight, by the mid-eighties he had realised that figural thinking, as a representative capacity of the human brain, needed to be treated as part of a ‘giant system of logical reasoning *networks*’.[43] If the science of simple systems was cybernetics, the science of giant systems was artificial intelligence.[44]

Around 1990, after the initial plan for the development of intelligent machines had been dropped, Qian turned his attention to *ren ji zhineng xitong* (人机智能系统)—human-machine intelligent systems as a pragmatic alternative to his optimistic vision of automated machines endowed with human intelligence founded upon noetic science, reasoning that if human intelligence could not be replicated, it must instead be integrated.[45]

Synthesis

In the decade that followed, Qian’s insights into human thinking, giant systems engineering, and computational systems based on human-machine interaction would come together in the new method of ’metasynthetic engineering’: a tool for effectuating ‘metasynthesis from the qualitative to the quantitative [从定性到定量综合集成]’ in the analysis of giant systems —probably the most important and under-theorised of all the ideas attributed to Qian.[46] The simplest model of metasynthetic engineering can be

understood to consist of three core components: human experts, machines, and knowledge systems. The central idea was to make all three work together to ‘solve the problems that arise in open complex giant systems’ and for which the methods used to analyse large systems were insufficient.[47] In practical terms, this meant a collective decision-making support system built upon expert knowledge and information infrastructure, designed to help manage large and complex yet ontologically disparate entities, such as a market or a war. In an early example provided by Qian, economists, management experts, and systems engineers would come together to work on a problem such as a price adjustment mechanism. A *leap* from the qualitative to the quantitative would occur when the three groups used their knowledge of their respective domains to determine how best to formalise (for example, by assigning environmental and control variables to a model) a possible solution to the problem as a *system*—by setting up a market monitor agency, for example. A simulation of the proposed solution would then be run, and the result tested for its effectiveness both by quantitative measures and by qualitative judgment (which is to say, by human experts), initiating another *leap*, this time from the formal model back to human judgment.[48] In this way, neither quantitative judgement nor qualitative judgement would be subordinated to the logic of the other—rather, the two would be integrated in such a way that their specificities would be taken into account.

It could be said that Qian’s metasynthetic engineering systems are the mature and implementable version of the social engineering systems of the previous decade—and, to risk ahistorical comparison for the sake of a concluding analogy, they are, in a sense, Cybersyn systems that survived and evolved to become a normalised and universalisable component of the late twentieth century and early twenty-first century computational paradigm.[49] Finally, it is fitting that Qian’s career-long negotiation of the qualitatively different demands of theory and practice, human thinking and machine thinking, would find its final form in a method for synthesising the two—and not just on the ‘micro’ level of the individual (which constitutes a giant system in its own right) but on the ‘macro’ level of social organisation, whose human complexities (and potentialities) Qian refused to simply subject to quantitative analysis, just as he refused to let the incomputable elements of human thought function as anything other than a spur to the continued refinement of intelligent machines.

Notes

1. [For a taste of debates over the influence of Soviet cybernetics, US strategic science, or European systems theory on Qian's work, see Wang Hongzhe and Jiang Yuan, 'Seeking for a Cybernetic Socialism: Qian Xuesen and the Transformation of Information Politics in Socialist China', *CAC Editorial* 1, 'Resurrecting Cybernetics': 249–59; and An, Bo, 'Modeled Modernity: A Critical History of Computing in the People's Republic of China, 1949 to 2000', PhD thesis, Yale University, 2022—ed.]
2. See Qian Xuesen (钱学森), *Collected Works of Qian Xuesen* (钱学森文集), ed. Gu Jihuan (顾吉环), Li Ming (李明), and Tu Yuanji (涂元季) (Beijing: National Defence Industry Press, 6 vols., 2012).
3. For these reasons among others, where there is a choice to be made I shall use the more distinctive terms for his Chinese concepts so as to avoid any confusion. For instance, I translate his 'thinking science' (思维科学) as 'noetic' not 'cognitive' science, even though both were suggested by Qian himself. See *Selected Letters of Qian Xuesen* (钱学森书信选)(Beijing: National Defence Industry Press, 2 vols., 2008), vol. 1 (Feb 1956–Dec 1991), 194.
4. See Wang, Ning, 'The Making of an Intellectual Hero: Chinese Narratives of Qian Xuesen,' *The China Quarterly* 206 (2011): 352–71.
5. For biographies, see I. Chang, *Thread of the Silkworm* (New York: Basic Books, 1996); Xi Qixin (奚启新), *Biography of Qian Xuesen* (钱学森传) (Beijing: People's Publishing House, 2011) and Ye Yongli (叶永烈), *Approaching Qian Xuesen* (走近钱学森), (Shanghai: Shanghai Jiao Tong University Press, 2009). For a two-volume detailed chronological biography, see Zheng Chengliang (郑成良总策划) and Zhang Xianmin (张现民) (eds.), *National Social Science Foundation Qian Xuesen Timeline Projec*t (国家社会科学基金项目钱学森年谱, 国家社会科学基金项目) (Beijing: Central Literature Press, 2015); Chang, *Thread of the Silkworm*, 63.
6. Chang, *Thread of the Silkworm*, 237.
7. As Qian wrote in 1979, when for him as for others in China, using scientific literature still largely involved the eyes (with few indexes available), the hands (copying without copy machines), and the feet

(transporting paper materials to scientists working in remote areas), in order to ‘modernise information science’, Chinese researchers needed new storage media and networked computer terminals to facilitate basic research so that they could focus more energy on mental and creative work. This, for Qian, would effectuate a great transformation in future scientific work. Qian Xuesen (钱学森), ‘The Modernisation of Intelligence Materials, Books, Documents, Archival Work, and its Impact [情报资料、图书、文献和档案工作的现代化及其影响]’, *Archives Communications* (档案学通讯) 5 (1979): 6–10.

8. *Toward New Horizons: Science, the Key to Air Supremacy* (A.A.F. Scientific Advisory Group, 1946).

9. Qian Xuesen (Hsue-shen Tsien), *Technical Intelligence Supplement: A Report of the AAF Scientific Advisory Group* (Wright Field, Ohio: Headquarters, Air Material Command, 1946).

10. Qian Xuesen (钱学森), ‘From Aircraft and Missiles to the Automation of Production Processes’ (从飞机、导弹说到生产过程的自动化), in *Collected Works*, vol. 1, 108–124.

11. Qian Xuesen (Hsue-shen Tsien), *Engineering Cybernetics* (New York: McGraw-Hill, 1954).

12. Ibid., vii.

13. Qian Xuesen, ‘From Aircraft and Missiles to the Automation of Production Processes’, in *Collected Works*, vol. 1, 109.

14. Ibid., 180.

15. Qian Xuesen (钱学森) et al., *On Systems Engineering* (论系统工程) (Changsha: Hunan Science and Technology Press, 1982), 2–27.

16. Ibid., 28–39.

17. Ibid., 32.

18. Ibid., 33.

19. Qian Xuesen (钱学森), ‘The Dialectics of Nature, the Science of Thinking, and Human Potential’ (‘自然辩证法、思维科学和人的潜力’), *Philosophical Studies* (哲学研究) 4 (1980): 7–13, 31.

20. Hubert Dreyfus’s well-known book *What Computers Can’t Do* (New York: Harper & Row, 1972) was yet to be translated at this time, but his ideas were known, and he was not the only one arguing for the categorical mistakes in AI claims. The Soviet debates on cybernetics, for instance, had taken place much earlier and were already being discussed in China in the 1960s.

21. The example of the Belle vs. Karpov chess game was intended to be indicative only of the contemporary difference between artificial intelligence and noetic science. Artificial intelligence, as Qian understood it, would eventually develop chess-playing machines powerful enough to beat the reigning world champion because it was rule-based and could theoretically be solved by symbolic or statistical models (later exemplified by Deep Blue an AlphaGo respectively).
22. Qian Xuesen, 'The Dialectics of Nature, the Science of Thinking, and Human Potential', 9–10.
23. At one point in his career, Qian would refer to a science for studying collective thinking patterns and mechanisms as 'social noetic studies'. *shehui siwei xue* (社会思维学). Qian Xuesen (钱学森), Zhejiang Science and Technology Association (浙江省科学技术协会) (1991), 63.
24. Qian Xuesen, 'The Dialectics of Nature, the Science of Thinking, and Human Potential', 12.
25. Note that prior to this period, Qian understood the specific content of AI research through second hand sources, mostly reference books—he cites a definition of artificial intelligence from a Computer Science encyclopaedia, for instance—and foreign popular science magazines such as *Scientific American*, in addition to Chinese translations and his internal scientific intelligence sources.
26. For a study of the 'qigong fever' of the period, see D. Palmer, *Qigong Fever: Body, Science, and Utopia in China* (New York: Columbia University Press, 2007).
27. Qian Xuesen (钱学森), *Human Science and the Development of Modern Science and Technology in a Long-Term Perspective* (人体科学与现代科技发展纵横观) (Beijing: People's Publishing House, 1996), 418.
28. Qian Xuesen (钱学森), *Selected Letters of Qian Xuesen* (钱学森书信选),(Beijing: National Defence Industry Press, 2 vols., 2008), vol. 2 (Nov 1992–Dec 1993), 318.
29. See his discussion of the difference between insight and inspiration in *Collected Works*, vol. 3, 323.
30. Qian Xuesen, *Human Science and the Development of Modern Science and Technology in a Long-Term Perspective*, 419.
31. Contemporary commentators would try to explicate the term by referring to literary, especially poetic, examples in the Chinese and Western

traditions. See Qian Xuesen (钱学森) et al. (eds.), *The Riddle of Inspiration* (灵感之谜) (Beijing: Beijing Normal University Press, 1986).
32. See, for instance Qian Xuesen, *Collected Works*, vol. 5, 293.
33. Qian Xuesen, *Selected Letters*, vol. 1, 320.
34. Ibid., 166.
35. Qian Xuesen (钱学森), *Introduction to Interstellar Navigation* (星际航行概论) (Beijing: China Aerospace Press, 2008), 166. This does not mean that Qian was an empiricist either. Indeed, he criticised Edward Feigenbaum for being an 'empiricist' who was blind to the need for a theoretical basis when building intelligent systems.
36. Qian Xuesen, *Collected Works*, vol. 5., 310. His close collaborators, including Wang Chengwei (Computing Expert Committee) and Dai Ruwei (戴汝为) (Intelligent Machines Sub-Committee), were among the committee members. Wei Hongsen (魏宏森) (ed.), *Centennial Celebration of Qian Xuesen's Connection with Tsinghua University* (钱学森与清华大学之情缘百年校庆) (Beijing: Tsinghua University Press, 2011), 126.
37. He suggested to researchers of object-oriented programming that they include figural thinking as well, thus demonstrating that he was not thinking in terms of AI and non-AI languages which centres around reasoning, but between figural and non-figural paradigms. Qian Xuesen, *Selected Letters*, vol.1, 320–334.
38. Wei Hongsen, *Centennial Celebration*, 111–12.
39. Qian Xuesen, *Collected Works*, vol. 5, 27.
40. Qian Xuesen (钱学森), *Qian Xuesen on Systems Science: An Anthology* (钱学森系统科学思想文选) (Beijing: China Aerospace Press, 2011), 70.
41. For a detailed account of the European influence on Qian, see An, Bo, 'Modeled Modernity'.
42. Qian Xuesen (钱学森), *Systems for the New Century* (创建系统学新世纪版) (Shanghai: Shanghai Jiao Tong University Press, 2007), 39.
43. Qian Xuesen, *Collected Works*, vol. 4, 167. Emphasis mine.
44. Qian Xuesen (钱学森) and Dai Ruwei (戴汝为), *On the Intersection of Dasein Wisdom Thinking Science, Literary Arts and Information Networks in Information Spac*e (论信息空间的大成智慧思维科学、文学艺术与信息网络的交融) (Shanghai: Shanghai Jiao Tong University, 2007), 139.
45. Qian Xuesen and Dai Ruwei, *On the Intersection*, 232.
46. Qian Xuesen, Yu Jingyuan, and Dai Ruwei, 'A New Scientific Discipline—The Study of Open Giant Complex Systems and Its

Methodology', *Chinese Journal of Engineering & Electronics* 4:2 (1993), 2–12: 5. Recall that Qian was retired and relied very much on his former collaborators including Dai Ruwei, Wang Shouyun, Wang Chengwei, and Yu Jingyuan through email exchanges.
47. Qian Xuesen, *Collected Works*, vol. 6, 274.
48. Qian Xuesen et al. 'A New Scientific Discipline', 5–6.
49. See E. Medina, *Cybernetic Revolutionaries: Technology and Politics in Allende's Chile* (Cambridge, MA: MIT Press, 2011).

You Mi

A Different Endpoint: Qian Xuesen and Cosmotechnics

How might we theorise computing beyond a Western perspective, and in particular through the lens of Chinese philosophy and history? This text takes the work of Qian Xuesen, a Chinese aerospace engineer with an interest in *qigong* (气功), as a critical starting point for approaching this question. Owing to his near canonisation in the Chinese history of technology, interpretations of Qian's ideas have often been superficial, and as a result his work on complex systems design has not been given enough attention. My exploration of his legacy began in 0, when I commissioned a film by artist Shi Qing for the Shanghai Biennale.[1] Exploring how Qian's work could inform our ideas about computation, the film took the form of a theory-fiction that speculated on the implementation of the hypothetical 'Yangtze River Computer' as a model for artificial intelligence in which qigong and what Qian called *renti kexue* (人体科学), or 'somatic science', are 'written back' into computation, uniting 'nature' and 'artifice', human and machine. Qian's project, as reimagined and elaborated on by Shi, makes an important contribution to discussions about a possible Chinese 'cosmotechnics', a framework for computation and technological development based in Chinese philosophy which has recently nourished animated debates around the prospect of alternative modernities.[2]

Qian Xuesen is a towering figure in the history of Chinese cybernetics. He spent the 1930s and '40s studying in the US, where he completed a doctorate at the California Institute of Technology under the supervision of mathematician, aerospace engineer, and physicist Theodore von Kármán, with whom he developed the Kármán-Tsien [Kármán-Qian] rule (for estimating compressibility effects of subsonic flow) in the field of fluid mechanics. In the mid-1950s, at the height of the Second Red Scare, Qian was forced to leave America and returned to China to oversee the development of the nuclear weapons, missile, and space programmes. In the decades that followed, he became the chief propagator of cybernetics in

China. His work advanced a particular branch of cybernetics (which he called ’engineering cybernetics’) that emphasised the importance of design principles in practical engineering and attempted to organise these principles into a holistic discipline united by fundamental concepts.

When Deng Xiaoping initiated the process of ‘Reform and Opening-up’ in 1978 he pulled the brake on heavy military industry and sought to decouple political mobilisation from technological development. As a result, revolutionary interests were no longer given priority. Instead, the aim was to kickstart China’s economy by ramping up consumer industries and allowing decentralised development driven by self-motivated local agents to gain traction. *‘Kexue shi di yi shengchanli’* (科学是第生产)—‘Science is the Primary Productive Force’—became the motto of the era. At this juncture in state economic policy, Qian continued to promote cybernetics by reframing it as the science of efficient management, and suggested, like his contemporary Gregory Bateson, that it be applied to the study of agriculture and ecology.

What Qian would eventually become most drawn to, however, were the *renti teyi gongneng* (人体特异功能) or ‘special capacities of the human body’ that he saw being manifested in *qigong* practices, which were booming in China during the 1980s and 90s (so much so that the period has become synonymous with ‘*qigong* fever’)[3] and which, alongside the idea that the human body is part of a vast communication system in which information can be transmitted via the invisible medium of *qi*, would come to underwrite his development of the new field of ’somatic science’. In October 1982, at the third plenary session of the Preparatory Committee for Human Science, Qian wrote a report entitled ‘Does this Breed a New Scientific Revolution?’.[4] In it, he declared that *qigong* and complementary investigations into supernatural somatic capabilities were a kind of human science worthy of in-depth study, and that they might lead to a new scientific revolution, the significance of which, he speculated, could be as profound for the twenty-first century as the scientific revolutions of quantum mechanics and relativity had been for the early twentieth century.

In the West, cyberneticians such as Warren McCulloch were interested in how to transpose mental processes into a kind of system that could be computed algorithmically. This reformulation of cognition in cybernetic terms as mere patterns that link thought to action marked the abandonment of ontological concerns with essences or the ‘being of things’, as expressed

in questions such as ‘What *is* x?’. In their place, an interest in function and the ability to predict future interactions, as expressed in questions like ‘What *does* x *do*?’, began to take precedence, displacing, in Qian’s eyes, the fundamental role of the human body in computation and the specificity of human thought. This shift nevertheless inaugurated a new field of methodologies and practices that would render computation so powerful that the models developed in the West have come to define the global paradigm today.[5]

Qian embarked on a different path. Not long after the publication of ‘Does this Breed a New Scientific Revolution?’, he began to propose research into what he called *kaifang de fuza ju xitong* (开放的复杂巨系统) or ‘open complex giant systems’, in which human experts, databases, various AI systems, and giant computers capable of performing billions of computations per second would come together to form an enormous hybrid human-machine intelligence. Open complex giant systems, he argued, should become the object of a whole new field of study in systems engineering, one which, because it differed so greatly from the cybernetics of the contemporary West, would open up an alternative route into computational research and development for China.[6] Far more broadly conceived than their Western counterparts, the systems Qian sought to create would be holistic: a ‘unification of natural science and social science’, synthesising technology with human thinking, collective organisational structures, and the ‘special capacities of the human body’.[7] Qian’s enduring wish was to integrate into computation something he called *linggan* (灵感) or ‘inspiration’—the human ability to make qualitative leaps in thinking—along with *xingxiang siwei* (形象思维) or ‘figural thinking’—the ability to solve complex problems very quickly without being aware of the logical steps one has taken to do so.[8] If a formal system based on logic alone fails to simulate human-level intelligence, he speculated, the pragmatic alternative was to create a symbiotic human-machine intelligence in which algorithmic logic and human creativity would combine without either one coming to dominate or instrumentalise the other.

Shi Qing’s *Qian Xuesen and the Yangtze River Computer* is a one-hour-long retro-futuristic science fiction film. It draws on real figures and events in the history of cybernetics in China while also adding fictional details, thus conjuring a speculative past and future for computation alongside Qian’s biography. The film is made up of archival footage, mostly in black

and white or desaturated colour. The narration switches between a matter-of-fact socialist-TV-anchor-like narrator, the protagonist Qian Xuesen, and an unnamed commentator who operates as a mouthpiece for the artist. News-report-style titles, key words, and important phrases are shown on screen in a red font using a typeface that resembles propaganda posters from the Cultural Revolution era, and the film's collage of large text blocks and images is reminiscent of the socialist *heibanbao* (黑板报), or 'blackboard newspaper', a grassroots form of reporting on current affairs, arts, and culture that uses blackboards placed in public spaces to disseminate information, serving as a tool for mass engagement.[9]

Vintage footage of Qian speaking at congresses and filling blackboards with mathematical equations is intercut with footage of avid workers and factory scenes. The visual and narrative style conjures up the feeling of watching *kexue jiaoyu yingpian* (科学教育影片)—'scientific education films', a genre that occupied a prominent part of Chinese screen time from the founding of the People's Republic well into the 1970s. Encompassing topics including hygiene, modern machinery and production methods, pesticides in agriculture, and communal organisation, they were often technical and painstakingly detailed while conveying a sense of collective progress. Sometimes targeting specific audience groups such as peasants, factory workers, mine workers, or students, these films are inextricably bound up with scientific education in line with Maoist theories of development.[10] To an unsuspecting viewer, *Qian Xuesen and the Yangtze River Computer* may appear to be a factual, scientific documentary. Narrated by stoic, objective voices, the film follows Qian's journey from the US back to China and documents his development of the 'Yangtze River Computer'—a supercomputer that integrates the collective organisational power of the masses to achieve optimal computational results.

In the film, the idea for the Yangtze River Computer comes to Qian during a visit to the construction site of the Shisanling Reservoir in Beijing in 1958 where he is struck by the orderly organisation of four hundred thousand voluntary workers, and remarks:

> This is like a computer only it is created out of the collaborative power of the physical masses! In contrast to the instrumental nature of Western computers, it is the consciousness and discipline of these workers which makes real computation possible.[11]

Qian wastes no time in translating his vision into a blueprint for a supercomputer, the core principle of which would be the conversion of 'the

organising energy of the natural world and human society into computational power'.[12] The government weighs in on this imaginative idea and the Central Military Commission sets up the 'Fifth Research Institute of the Ministry of Defence', a covert unit, with Qian as its director. It is decided that the supercomputer will be located near the Ma'anshan Iron and Steel Plant on the lower reaches of the Yangtze River where it can profit from both the organisational discipline of the steel workers and the hydraulic power of the Yangtze River, from which it inherits its name.

The ontological premise of the Yangtze River Computer as a 'computer of and for the masses' is an extrapolation from the mass movements in which the Chinese cybernetics program was rooted. Chinese cybernetics did not follow the path of Soviet cybernetics, which was largely controlled by a technocratic elite and thus presumed a high level of centralisation. The Soviet cybernetics programme ultimately failed as a result of self-interest and fierce competition among the numerous research institutes, technocrats, and specialists involved.[13] However, China did not have a technocratic class like the Soviet Union, and during the periods of the Great Leap Forward and Cultural Revolution especially, Chinese scientists worked on cybernetics, mathematics, and data science in person at local production sites such as villages and factories alongside peasants and workers, with the goal of effectuating mass political mobilisation through technology. Until the end of the 1970s, the Chinese approach to technological development was characterised by equal emphasis on high-level scientific projects and the grassroots technological politics of the 'mass line'.[14]

When the Yangtze River Computer faces a bottleneck in productivity owing to issues encountered, firstly with the diversity in levels of knowledge, training, and experience of the individuals (researchers, intellectuals, skilled workers, and students) comprising its computational substrate, and secondly with attempts to scale up its computational power, a Chinese delegate from the Fifth Research Institute is shown travelling to Salvador Allende's Chile to share information and, under instructions from Qian, to garner inspiration from Project Cybersyn. Designed by American cybernetician Stafford Beer, Cybersyn aimed to support the management of the national economy by processing production input and output data in real time whilst involving workers in management. It is noted that, like the Yangtze River Computer, Cybersyn 'places the human factor at the centre' and 'seeks to create a symbiotic system of diversity and autonomy between

computers and organised networks of workers'. At the same time, the flow of information through the system no longer depends on a traditional hierarchical bureaucracy but instead on a structure in which 'the public and experts work together in a centrally negotiated way'.[15]

The film also identifies similarities between Cybersyn and the Angang Constitution (*Angang Xianfa* [鞍钢宪法]) of 1960—a set of management axioms derived from the practical experience of workers at the Anshan Iron and Steel Company.[16] Reflecting on the disparities present among the human components of Yangtze River Computer, Qian wonders:

> Might the Angang Constitution be a solution to this difficulty? Should our researchers and leaders go down and participate in the computational work while more workers join the leadership group and participate in management? Should all rules and regulations found to be unreasonable be reformed? And should we implement an integration of workers, leading cadres, managers, and technicians? In the future, these will become the new organisational principles, the core algorithms of the project.[17]

As Shi's semi-fictional historical account reaches the era of Reform and Opening-up, Qian is forced to surrender the project to a wider scientific and industrial application in line with Deng's policy. However, he does not give up on the collectivist organisational principles inherited from the Angang Constitution. The result of this compromise is a parallel computing process involving both modern computer hardware and storage systems and the collective labour of the computing masses. Here Qian is depicted critiquing the American model of the personal computer:

> I think the defect of the American computer is that, to put it bluntly, it is an object without a subject. If the Yangtze River Computer were to replicate it or to lose its original organisational structure, this would be equivalent to replacing the subject with an object and production with consumption. A total separation of the computational subject from the organisational mechanism would only transform the Yangtze River Computer into a sort of 'zombie computer'.[18]

In a final speculative leap, Shi traces Qian's deep interest in *qigong* and the 'special capacities of the human body' back to the work of his colleague at the California Institute of Technology, Jack Parsons, a rocket engineer under von Kármán who 'secretly participated in a communist discussion group with Qian' and practiced black magic, which he believed was 'theoretically intertwined' with quantum physics.[19] Amidst real reports from the provinces of children and workers able to 'read' Chinese characters without looking at them, see through barriers, or practice telekinesis, Shi's Qian Xuesen surmises that these 'special capacities' are the result of 'energy deviations produced by quantum leaps under unconventional

conditions' occurring during secret government tests that took place during Mao's Third Front Campaign at sites close to the operational bases of the Yangtze River Computer.[20]

The film concludes by telling us that although Qian passed away in 2009, his theories, alongside China's real world-leading advances in quantum computing, have, by connecting somatic science, quantum computing, and the labour of the computing masses, led to the construction of a new supercomputer in which, following the original organisational principles of the Yangtze River Computer, the computing masses are not instrumentalised by a centralised control unit but incorporated into the computational process as its active components, evolving with and being emancipated by the technology rather than being alienated and enslaved by it:

> The supercomputer is not a machine, but a symbiotic aggregate with computing as its organising technology. This 'cyborg', which includes all things in nature and all people, has another name: 'common cause'![21]

*

Through this mix of real and fictitious events staging an alternative history of computation, the narrative presented in *Qian Xuesen and the Yangtze River Computer* complicates the way in which futurism is often imagined in relation to China. It may be called a 'sinofuturist' film by some accounts, and yet it remains importantly distinct from Sinofuturism, an artistic genre that turbocharges technological imaginaries of China. With a touch of cyberpunk noir, Sinofuturism typically foregrounds the speed and velocity of technological innovation, ultra-efficiency as a mantra of statecraft, and bottomless obedient crowds as an abundantly available resource for technological development.[22] According to Gabriele de Seta, Sinofuturism operates with an aesthetic repertoire that is legible to an audience both inside and outside of China, evoking a sense of fascination, anxiety, and national or cultural pride. However, it reproduces techno-orientalist tropes and narratives premised on a 'denial of coevalness'.[23] The idea that China exists on a different timeline to the rest of the world, one marked by mechanical mass mobilisation, unites Sinofuturist imaginaries with the twentieth-century Western Left's fascination with communist states and their supposed anti-individualism.[24] It is also present in some local imaginaries, such as the infamous (arguably techno-orientalist) analogue

computation scene in the blockbuster Chinese science-fiction trilogy *Remembrance of Earth's Past* by Liu Cixin. In a simulated world, Qin Shi Huang, the First Emperor of China in the third century bce, orders his thirty-million-man army to enter a formation that allows it to function as a computer. Each soldier signals a binary value with a white flag for '0' and a black flag for '1', collectively forming the circuitry of an enormous motherboard.[25] In typical Qin Shi Huang fashion, the penalty for a soldier making a mistake and causing the computation to abort is execution.

These imaginaries of the human masses as a giant machine or robotic horde only replicate the assumptions behind the trope of the 'yellow peril', if not also fundamentally misunderstanding the paradoxical relation between living labour and dead labour.[26] The fiction of the Yangtze River Computer presents an alternative: it debunks both the universalist computer and the fetishistic, localist Sinofuturist supercomputer. In their place, it asks what other kinds of computation might be possible and desirable. Playing with truth and speculation, the film allows the intensities of the present to emerge rather than locking them into the models of a barbaric past or a fully automated machinic future. Qian Xuesen's work—especially his interest in collective *qigong* practices—suggests an alternative conceptualisation of the masses, of collectivity, and of computation. Rather than a throng of mindless workers at the disposal of an algorithmic regime, how might we think of a human mass that has *agency?* This is, in some ways, a philosophical problem. But Qian's speculations allow us to venture beyond the dominant Western cultural approach to computational paradigms, lending themselves instead to a 'cosmotechnical' approach. Shi Qing's vision of Qian Xuesen focuses on material constellations of technology, neither advocating for technological dominion over humanity nor for a scaling back of technology in the name of the human. The Yangtze River Computer is not a human-machine assemblage that leads to subsumption, servitude, and alienation, but it also resists the model of a grounded, 'local', cosmological episteme defined by practices that are small-scale, analogue, and communal. In the film, technology does not have to be voided of speculative dimensions to be viable. Qian's story shows how these practices can be lived and translated into concrete design.

In a conversation with philosopher (and ideologue of Eurasianism) Aleksandr Dugin, Yuk Hui suggests that if the Western model of technological singularity or 'AI takeover' marks the moment when human

Dasein is destroyed, then instead of viewing ‘the singularity’ as the synchronised endpoint of history, we should diversify the notion of ‘singularity’ to take account of different endpoints.[27] Computational pluriverses can be conceived of by working backwards from the kind of endpoints desired by different communities. It would be possible, for example, look to traditional wisdom such as the Japanese *kami* (deities that are also forces of nature) to counteract the alienation caused by AI and enact a process of rehumanisation.[28] That said, a solipsistic retreat into a prepackaged cosmology threatens to be at best endearing but ineffectual, at worst a cold operationalisation of traditional knowledge. Benjamin Bratton warns of the pitfalls of this approach, for when divergence becomes a value in and of itself, we lose sight of technological viability. Importantly, the relation between convergent and divergent instrumentality and convergent and divergent epistemology (or cosmology) is not the same. As Bratton notes in ‘The Stack at the Edge of Planetarity: Convergence, Divergence and War’, ‘[w]e can imagine Stacks which are technologically diverse but cosmologically monocultural and we can imagine Stacks which are technologically isomorphic but animated by diverse cultural imaginaries (and presently, we have them)’.[29] For those of us with a specific interest in alternative paradigms of computation, the question that *Qian Xuesen and the Yangtze River Computer* brings to the fore is: Could we characterise Qian’s design for a large complex system, in which human styles of thought and problem solving (such as *linggan* and *xingxiang siwei*) are reintroduced alongside somatic science, as cosmotechnical?’ If the Yangtze River Computer is the endpoint, where do we begin?

The work of Qian Xuesan diverges from today’s dominant technological systems at both the instrumental and the epistemological level. However speculative Shi’s image of the Yangtze River Computer, Qian clearly had an interest in viable reconfigurations of technological systems design. Given that Chinese computer scientists are still deeply influenced by his theories and are indeed working on contemporary implementations of Qian’s legacy, this ‘could have been’ of computation is significant in more than just a symbolic way. There is a practical value to be discovered in rearranging the material assemblages of human-machine computational systems. For Qian, human513 *linggan* and the ’special capacities of the human body’ are never presented as incomputable and therefore unoperationalisable. Instead, he was convinced that a time would come when research on their mechanisms

would be complete and they could be fully incorporated into computation, completely transforming the contemporary, reductionist515 computational paradigm. In a similar movement of thought, can we not ask, what would the social use of a Yangtze River Computer look like in our time? And if the social(ist) reclamation of algorithmic organisation today involves mobilising the emergence of complexity against complexity,[30] can the speculative idea of the 'calculating masses' not be repurposed as a dynamic organisational structure to meet the demands of other kinds of complexities, specifically those that confront us in this current climate of geopolitical turmoil?

Notes

1. Shi Qing (石青), *Qian Xuesen and the Yangtze River Computer* (钱学森和长江计算机), single-channel video, 2021. The film was commissioned for the *13th Shanghai Biennale: Bodies of Water*, 2021.
2. Coined by the philosopher Yuk Hui, 'cosmotechnics' recognises the unification of cosmic and moral orders through technical activities as a means of overcoming the conceptual dualism of technics and nature. For Hui, technics or tools—*qi* (器)—are to be understood as belonging to the cosmic order—*Dao* (道)—not as something opposed to it. Hui also takes issue with the idea, which he sees as embedded in the very notion of technics, that there is a single, universal nature, whose telos can be progressively revealed via technology. Instead, he argues that such conceptions themselves belong to the specific political order of Western modernity. Consequently, he advocates for a great plurality of technics—or cosmotechnics—to reinvigorate cosmopolitics beyond this stale Western technological monoculture. See Hui, Yuk, *The Question Concerning Technology in China* (Falmouth: Urbanomic, 2016) and Hui, Yuk, 'Cosmotechnics as Cosmopolitics', *e-flux* 86 (2017).
3. D.A. Palmer, *Qigong Fever: Body, Science and Utopia in China* (New York: Columbia University Press, 2007).
4. Qian Xusesen (钱学森), 'Considering Science Popularisation as a Great Strategic Task' (把科普工作当作一项伟大的战略任务来抓), *Creating Popular Science* (科普创作) 3 (1980), in *Collected Works of Qian Xuesen*

(钱学森文集), ed. Gu Jihuan (顾吉环), Li Ming (李明), and Tu Yuanji (涂元季) (Beijing: National Defence Industry Press, 6 vols., 2012), vol. 2, 349.
5. O. Halpern, *Beautiful Data: A History of Vision and Reason since 1945* (Durham, NC: Duke University Press, 2014), see especially Chapter 3.
6. Qian Xuesen, Yu Jingyuan, and Dai Ruwei, 'A New Discipline of Science [A New Scientific Discipline]—The Study of Open Complex Giant Systems and Its Methodology', *Chinese Journal of Engineering & Electronics* 4:2 (1993), 2–12. With his open complex giant system, Qian foresaw the developments in big data, the internet of things, and smart-city related technologies that are familiar to us today, but for Qian they were inflected with a more human-centred approach.
7. Ibid., 11; Dai Ruwei (戴汝为), 'From *Engineering Cybernetics* to Metasynthetic Engineering: Commemorating the 50th Anniversary of Qian Xuesen's Return to China' (从工程控制论到综合集成研讨厅体系 — 纪念钱学森归国50周年), *Complex Systems and Complexity Science* (复杂系统与复杂性科学) 3:2 (2006).
8. For a detailed discussion of the specific ways in which Qian understood and used these terms, see An, Bo, 'Between the Human and the Machinic: Qian Xuesen and AI', in this volume.
9. The use of socialist aesthetics and a documentary style has been a hallmark of the artist's recent works, which critically revisit the legacy of socialism in China. In *Opposite the Zoo is the Planetarium* (2018), Shi traces parallels with Russian cosmism in China, and connects the cosmist project of resurrection with the making of a new socialist subject. *Fungus, Antimatter, and Fluid Urbanism* (2019) speculates on how the secret military industrial complex established during the socialist period left a lasting mark on the city of Chongqing. Such allusions to socialist aesthetics are also prevalent in the contemporary art scene in China and have reenergised interpretations of works by artists related to the Political Pop movement. See P. Wang, *The Future History of Contemporary Chinese Art* (Minneapolis: University of Minnesota Press, 2021).
10. See M.D. Johnson, 'The Science Education Film: Cinematizing Technocracy and Internationalizing Development', *Journal of Chinese Cinemas* 5:1 (2011): 31–53.
11. Shi, *Qian Xuesen and the Yangtze River Computer.* Translation modified.
12. Ibid.

13. See B. Peters, *How Not to Network a Nation: The Uneasy History of the Soviet Internet* (Cambridge, MA: MIT Press, 2016).
14. See Wang Hongzhe (王洪喆), ‘The Political Turn in Chinese and Soviet Cybernetics and Socialist Information Technology: A Historical Comparison’ (中苏控制论革命与社会主义信息传播技术政治的转型：一项基于文献的历史比较), *Remapping* (区域) 1 (2016).
15. Shi, *Qian Xuesen and the Yangtze River Computer.* Translation modified.
16. The Angang Constitution was designed to counter the outsized power of technocrats, which was seen to be affecting the operation of Soviet enterprises at the time, and to break the impasse in stimulating technological innovation in both China and the Soviet bloc. The larger context was the threat of capitalist restoration, which the constitution aimed to curb through mass political mobilisation. See Y.Y. Kueh, *China’s New Industrialization Strategy: Was Chairman Mao Really Necessary*? (Cheltenham: Edward Elgar, 2008). In the Cybersyn system there are two ways to tame complexity. One is by way of rules, so as to make internal behaviour within the system more uniform, and the other is by granting autonomy and power to individual components so that they can re-engineer the underlying organisational structure in the face of emerging complexities. A comparative study of the Angang Constitution and Cybersyn certainly merits further treatment from the fields of media theory and economics. See S. Beer, *Designing Freedom* (Toronto: CBC Publications, 1974).
17. Shi, *Qian Xuesen and the Yangtze River Computer*. Translation modified.
18. Ibid. Translation modified.
19. Ibid.
20. Ibid. In the film Qian proposes that, at the close of the period of *qigong fever*, state-backed claims of *qigong* being a delusion can only arise from viewing the practice through the lens of ‘information individualism’—abstracting the human body out of the system it is embedded in and treating it as an ‘independent computer’.
21. Ibid. Translation modified.
22. See L. Lek, *Sinofuturism* (中华未来主义) *(1839–2046 AD)*, single-channel video, 6, and V.L. Conn, ‘Sinofuturism and Chinese Science

Fiction: An Introduction to the Alternative Sinofuturisms (中华未来主义) Special Issue', *SFRA Review* 50, no. 2–3 (2020): 66–70.
23. G. de Seta, 'Sinofuturism as Inverse Orientalism: China's Future and the Denial of Coevalness', *SFRA Review* 50, no. 2–3 (2020): 86–94.
24. See for example B. Noys, *Malign Velocities: Accelerationism and Capitalism* (Winchester: Zero, 2014).
25. Liu Cixin, *The Three-Body Problem*, tr. K. Liu (London: Head of Zeus, 2016), 223–43. Importantly, the individual components do not themselves need to be intelligent:

> 'Instead of mathematicians, we'll use common laborers. But we need many of them, at least thirty million. We'll do mathematics using human wave tactics.'
>
> 'Common laborers? Thirty million?' Wang was amazed. 'But if I recall correctly, this is an age when ninety percent of the population is illiterate. Yet you want to find thirty million people who understand calculus?'
>
> 'Have you heard the joke about the Army of Sichuan?' Von Neumann took out a thick cigar, bit off the end, and lit it. 'Some soldiers were being drilled, but because they had no education, they couldn't even follow the drill instructor's simple orders to march LEFT-RIGHT-LEFT. So the instructor came up with a solution: He had every soldier wear a straw shoe on the left foot and a cloth shoe on the right. When they marched he shouted … STRAW-CLOTH-STRAW-CLOTH…. That's the kind of soldier we need. Except we need thirty million of them.' (225–6.)

26. 'Communist accelerationism', according to Benjamin Noys, tries to counteract the capitalist dynamic in which living labour (i.e. human workers) is squeezed out by 'dead labour' (i.e. machinery) leading to a lack of employment for human workers. It does so by reversing the dependence of living labour on dead labour and by fusing the human body of living labour with the machines of dead labour, thereby forming living machines. (Noys, *Malign Velocities: Accelerationism and Capitalism*, Chapter 2, 'Leaps! Leaps! Leaps!: Communist Accelerationism'). Notwithstanding the continuities and ruptures therein, defining living labour as human workers and dead labour as machines simplifies the dynamism of capitalism. In a grander scheme that takes cycles of investment and the global market into consideration, dead labour can be analysed in Marxist terms as fixed capital, with investments in, for example, better machines and infrastructures realising their value over time and, along the way, enabling

more production for a greater base of labour power. Hence, paradoxically, ‘dead’ labour leads to more ‘living’ labour, or the massification of labour and consumption at a global scale. See D. Harvey, ‘Rate and Mass: Perspectives from the *Grundrisse*’, *New Left Review* (July/August 2021).
27. ‘“The singularity” has been commonly perceived as an endpoint. All civilisations and cultures—Russia, China—we are synchronised by this endpoint towards “the singularity”. […] [But] instead of converging and being synchronised to this singularity, is it possible that we will open it [up]—as a multiplicity, as multiple histories? […] It seems to me that we need to diverge [from] this convergence—this to me is even more speculative than speculating on convergence on one point […] and this is also the core of the argument of multiple cosmotechnics.’ Hui, Yuk and A. Dugin, ‘Technical Issues’, S*reda* (podcast), 2020, link no longer available.
28. J. Vallverdú, ‘The Eastern Construction of the Artificial Mind’, *Enrahonar, Quaderns de Filosofia* 47 (2011): 171–85: 177. Cosmotechnics, for Hui, is not necessarily related to computation but works as a sort of analogy for it, with Hui describing Chinese medical practice as cosmotechnical, expanding the definition of technology to encompass these ‘minor’ practices.
29. B. Bratton, ‘The Stack at the Edge of Planetarity: Convergence, Divergence and War’, in *Vertical Atlas*, ed. L. Dellanoce, A. Khalaf, K. Kuitenbrouwer, N. Nyabola, R. Roukens, A. Steiner, and M. You (Arnhem: ArtEZ Press, 2022), 276.
30. See D. Saros, *Information Technology and Socialist Construction: The End of Capital and the Transition to Socialism* (Abingdon and New York: Routledge, 2014), and E. Morozov, ‘Digital Socialism? The Calculation Debate in the Age of Big Data’, *New Left Review* 116/117 (2019): 33–67.

Wang Hongzhe

Five Moments in the History of Chinese Cybernetics

Translated by Allen Young

China converts its converters.

NORBERT WIENER[1]

I

In September 1935, the Electrical Engineering Department at Tsinghua University welcomed a visiting research professor from the United States. His name was Norbert Wiener, and he had come to spend a year in China at the invitation of Hiong King-lai and Lee Yuk-wing, with the support of the university president, Mei Yi-chi. Wiener lectured in the departments of Mathematics and Electrical Engineering on Fourier integrals and series, and he worked with Lee on an initial attempt to build an 'analogy-computing machine'. He also developed a keen interest in the culture, society, and living conditions of this—to him—mysterious country.[2]

Wiener enrolled his two young daughters in the American school at Yenching University, where they began learning Chinese. Outside of class, as he recalled in his memoir *I Am a Mathematician*, he played chess, go, and five-in-a-row with his colleagues. He and his wife took rickshaws into the city and felt 'ashamed at being pulled along by the strength of another man'. He found Peiping (as Beijing was then known) to be 'a mixture of glamor and squalor', noting the Japanese presence and observing how students 'marched into the city to protest the coming of the Japanese and the supineness of the Chinese'. Wiener was struck by something nearly all his Chinese friends shared: 'that love of the whole world rather than any specific humanity, which is so characteristic of Buddhism'. And despite the 'broad and durable companionship between many of the Westerners and

many of the Chinese', he observed that 'the Chinese all felt strongly that they should be masters in their own house'.[3]

Looking back, Wiener saw the year he spent travelling with his family in China and around the world as a watershed moment in his life. 'If I were to take any specific boundary point in my career as a journeyman in science and as in some degree an independent master of the craft', he wrote in his memoir, 'I should pick out 1935, the year of my China trip, as that point'. That year gave his children 'the enormous moral advantage of seeing the world as a whole and not merely as an interplay of master races and servant races'.[4] Later on, after he returned home, Wiener actively took part in fundraising efforts to support the Chinese people's war of resistance against Japan.

II

In 1955, missile expert Qian Xuesen (also known as Hsueh-shen Tsien) left the United States on a steamship bound for China, where he would found the new Chinese government's missile and rocket industry.

In the late 1940s, Qian had been a member of the US Air Force Scientific Advisory Group led by Theodore von Kármán, and had taken part in confidential intelligence and technology planning. This introduction to the field of military operations research led him to begin thinking about cybernetics' potential expansion, after the war, into the social sciences. No doubt this may have been part of the reason that the McCarthyists accusing him of communist sympathies sought to keep him in the US: not only did they worry that Qian would take his missile expertise to the Chinese or the Soviets, they also feared losing a strategic thinker for the Cold War.

In the four years he spent under partial house arrest, Qian wrote *Engineering Cybernetics*, which he published in 1954. Building off Wiener's earlier *Cybernetics*, Qian's book was quickly translated into several languages and became one of the most frequently cited foundational theoretical texts on automation. In 1955, after negotiations with China, the US agreed to let Qian return home in exchange for eleven pilots captured by the Korean People's Army.

On the steamship home, Qian discussed the idea of combining operations research and the socialist planned economy with Xu Guozhi, another returning Chinese mathematician. Qian believed the war had allowed the

US to make rapid advances in military logistics through the extensive application of mathematics and operations knowledge. China, as a socialist country with a highly planned society, could make ample use of operations research and cybernetics, which he believed could be more fully developed under the conditions of socialism.

One of Qian's initial tasks upon his return was to set up China's first operations research group. He did this with Xu in 1956, as part of the Mechanics Institute (which Qian also set up) at the China Academy of Sciences. Control theory became one of the group's four major research areas. In 1958, Hua Luogeng, a mathematician who had attended Wiener's lectures at Tsinghua, set up an operations research laboratory at the Mathematics Institute at the CAS, which absorbed Qian's group in 1960. That same year, Qian founded China's first military operations research centre—the Combat Research Department at the Fifth Academy of the Ministry of National Defence.

The Chinese edition of *Engineering Cybernetics* appeared in 1958. Four years later, to respond to national defence needs, Qian recruited Guan Zhaozhi, and also Song Jian, who had just returned from studying cybernetics in the Soviet Union, to set up China's first cybernetics laboratory in 1962, in conjunction with various industrial departments.

The work of Qian Xuesen and other returning scientists can be called the origin of Chinese cybernetics. In keeping with Qian's principles of pragmatism, the cybernetics laboratory instituted a '3-3 system', with a third of its scholars drawn from the sciences, a third from engineering, and a third from the social sciences. It was the first institution to have scientists and engineers working side by side with social scientists. With his experience in military strategy offices in the US Air Force, Qian was well aware that collaboration of social scientists and mathematicians was the methodological underpinning for this integrated strategic science. As early as the mid-1950s, when he had just returned, Qian proposed applying quantitative methods to the analysis of China's economy and society. His ideas about quantification and master plan simulations were as advanced as Cold War thinking on the social sciences and social cybernetics in the US and the USSR, if not more so.[5]

III

In his 1956 speech ‘On the Ten Major Relationships’, Mao Zedong wrote: ‘We must not follow the example of the Soviet Union in concentrating everything in the hands of the central authorities, shackling the local authorities and denying them the right to independent action […]. In short, if we are to promote socialist construction, we must bring the initiative of the local authorities into play.’[6]

Even more than national independence, Mao’s vision of revolutionary development sought autonomy at every level of society. One key precondition for achieving autonomy was the widespread use of electronic technology—an electronic revolution to develop the initiative and creativity of the masses. Hence, in the 1960s, Mao’s right hand man Chen Boda proposed the notion of ‘electrocentrism’, which meant that the electronic industry should be a development priority not only in the field of cutting-edge technology, but throughout the country’s economy. The belief that only large enterprises could produce electronics had to be eradicated: the country needed to ‘smash electronic mysticism’ and burst into the electronic age with the methods of a people’s war.

Although this fact is often overlooked in contemporary histories, the Cultural Revolution went hand in hand with a ’Cyber Great Leap Forward’. As early as 1968, the output of local electronic industries had already surpassed that of central enterprises (¥825 million against ¥494, respectively). It extended not only to provinces with a good industrial infrastructure but even to marginal regions with no experience in electronics production, such as Qinghai, Ningxia, and Tibet. To guard against a full-scale invasion from the north by the Soviet Union, Beijing’s electronics manufacturers helped set up a number of ‘third-tier enterprises’ in the centre and west of the country. Provinces such as Shaanxi, Shanxi, Gansu, and Guizhou acquired complete industrial capabilities in military electronics and their supporting services. In 1970, China had two and a half times as many electronics factories as in the previous year, and twenty times more than in 1965 before the start of the Cultural Revolution.[7]

In Mao’s thinking, even if China were to lose most of its territory in the north, including the capital, it would still be able to rally forces from the hinterland and wage an all-out ‘electronic’ guerrilla war. The Great Leap Forward in local electronics industries makes it clear that China’s development goal was to advance toward a society-wide ‘cyberneticisation’ at every level. What the people needed was not a colossal electronic city

capable of satisfying production needs, but ‘electronic Paris communes’ spread out over the entire territory.

Along with the ‘Cyber Great Leap Forward’ came ‘cyber mass movements’ based on ‘three-in-one’ groups. Across the country, the neighbourhood factories and textile mills run by women began to produce electronic components and even computers. For a time during the 1970s, textile workers in these small collective factories became the primary force in computer manufacturing. This little known history has its parallels in California’s ‘garage culture’ and counterculture.

After China and the US reestablished relations in 1972, the two countries exchanged science and technology delegations. One factory that the Chinese arranged for American computer experts to visit, the Yangtze River Radio Factory, had been converted from producing door handles to producing electronic computers. As the 1973 delegation report noted:

> The new factories we visited were varied both in size and in manufacturing technology. At one extreme is the factory associated with the Shanghai Computing Research Institute. One of China’s numerous neighbourhood factories, it is staffed by people in the neighbourhood and is therefore small […]. The American visitor is struck by its good order; generous spacing of work tables; extreme simplicity of facilities; the quiet movement of engineers and other workers in the course of their hand work, study, and discussion; sunshine; and incense.[8]

In the winter of 1974, this same factory organised a symposium entitled ‘Learning the Dialectics of Nature by Doing: Independence, Self-Reliance —Toward the Development of Electronic Computers’. Some years later, a round table on ‘people’s computers’ was translated into English and included in a reader edited by communications scholar Armand Mattelart, who cited it as the world’s first instance of a democratic political practice in the field of information technology.[9] In the 1970s, the red cyber revolution had reached its crescendo, even as it began drawing to a close.

IV

In 1979, at the fifth National People’s Congress, Qian put forward a proposal for ‘Establishing an Archive System and Electronic Computer Retrieval Network for People’s Ideas and Suggestions’, which contained his embryonic thinking on cybernetics and big data. Throughout the 1980s, Qian devoted himself to expanding the mathematical methods and management experience he had used in missile development to the fields of social cybernetics and systems theory. Legal cybernetics was an important

part of this—and here we discover a thread linking Qian and cybernetics to the poet Hai Zi (Zha Haisheng).

Known today as one of China's major modern poets, Hai Zi had another significant identity as a founding member of the legal systems theory group at China University of Political Science and Law. He began working as an editor at the university newspaper in 1984, and in September of that year he joined the Philosophy Department, where he taught courses in cybernetics, systems theory, and aesthetics.

Inspired by Qian's thought on legal systems engineering, Hai Zi and his friends founded the Legal Systems Research Association in April 1985. Hai Zi served as deputy secretary general and had a hand in organising the First National Scientific Symposium on Legal Systems. Qian was invited to attend and gave a keynote address on 'Modern Science and Technology and the Construction of a Legal System'. He advocated a role for systems theory in building a legal system and proposed replacing forensic investigations with new methods in information and biotechnology.

On the Foucauldian question of a 'surveillance society', Qian made his techno-optimism clear in 1984, in an article he co-authored with Wu Shihuan:

> The tools and principles used by the bourgeoisie can also be used by the proletariat. We must persist in class analysis, but on a dialectical basis. We must no longer make the mistake of indiscriminately assigning class labels. […] We must stand on the side of Marxism, using [technology] as a tool to serve the construction of modern socialism. Far from shackling the hands of the proletariat and the broad masses of people, it will extend the proletariat's reach, making it more flexible, more agile, more effective.[10]

It was at this conference that Hai Zi met Chang Yuan, a mysterious figure researching legal systems theory and so-called 'special capacities of the human body' or '*renti teyi gongneng*' (人体特意功能), i.e., paranormal abilities. (During the same period, Qian was one of the primary promoters of the Special Capacities of the Human Body Research Group.) Chang was a devoted follower of Qian's systems theory, and he had previously published Hai Zi's article 'On State Formation and the Function of the Law Viewed from the Perspective of Catastrophe Theory' in his journal *Tantao* [*Inquiries*]. In July 1985, he took up a teaching post at the Central Leadership Institute of Politics and Law and moved into the same faculty housing compound as the poet. Chang's wife, Sun Ge, as a 'gifted' person with paranormal abilities, was specially recruited by the Ministry of Geology and Mineral Resources to attend China Geosciences University in

Wuhan in 1988. In Changping, Hai Zi, Chang, and others formed a lively group discussing systems theory, sciences of the body, and *qigong*. As Chang later recalled, they would often drink and debate weighty issues such as national development. Around the same time, Hai Zi followed Chang and Sun in taking up *qigong*.

On March 26, 1989, at the age of 25, Hai Zi lay down on the train tracks near Shanhaiguan and took his own life. The circumstances surrounding his death have always been murky, and in one of his last writings, he pointed an accusing finger at Chang and Sun:

> March 24, 1989: Tonight I'm acutely aware that it's Chang Yuan and Sun Ge, with their Daoist witchcraft, who are making me hear voices. Most of the voices come from them. If I'm struck down by sudden death, schizophrenia, or suicide, the blame lies with them. They must be held criminally responsible.[11]

In an interview more than twenty years later, Chang admitted that the young poet glowed with energy and could see things that ordinary people could not. Hai Zi's death is inextricably bound up with the rise of systems theory, paranormal research, and *qigong*—forgotten chapters in the shifting social thought and culture of 1980s China that are now footnotes in the history of Chinese cybernetics.

V

The China moment in the history of cybernetics is shot through with a dialectic of labour and machines. In the revolutionary mindset of twentieth-century China, workers were always a more important strategic resource than technological wares. In the period from the 1950s to the 1970s, there is a discourse about the 'socialist line on technology' in many texts about computing, both popular science and science fiction. In this discourse, the goal of technical innovation was not to replace human labour with machine production, but to forge 'new people', *xinren* (新人), who could build and operate machines to serve the new society. Here the possibilities of technology were not disputed or downplayed but rather amplified. A new cybernetic loop took shape: machines would create 'new people', who would in turn recreate and modify the new society and its new adaptive technologies. On the one hand, this answered the Chinese revolution's modernising call to 'combat feudal superstition', while on the other, by

pushing to ‘democratise’ participation in technology, it answered the revolutionary call to destroy the monopoly of knowledge.

On a final note, this historical context casts a new light on Liu Cixin’s early short story ‘Sun of China’ (2002).[12] Perhaps even more than *The Wandering Earth* or *The Three-Body Problem*, this story reflects certain ‘Chinese characteristics’ distinct from the West in technology-related issues of politics and labour. Shuiwa, a boy from the countryside with only an elementary school education, is recruited for a mission to space and lifted out of his lowly status in the market society. Here technopolitics and technopoetics are briefly united in an echo of Wiener’s concern for ‘the human uses of human beings’.

We nevertheless see difficulties in this attempt to bridge past and present. In the world’s largest factory cluster, on the Pearl River delta, automation is gradually replacing human labour. Workers eliminated from manufacturing jobs can only join the gig economy and find work in food delivery, ride sharing, or gaming apps. They are no longer ‘new people’ but have sunk to an abject ‘bare life’ in the cybernetic loop of capitalism.

Against the backdrop of China’s rise as the world’s second largest economy and the trade and technology war with the United States, our ability to return to forgotten moments in the history of Chinese cybernetics and reactivate its circuits of labour may hold a key to the future of this ancient civilisation.

Notes

1. N. Wiener, *I Am a Mathematician: The Later Life of a Prodigy* (Cambridge, MA: MIT Press, 1956), 197.
2. I have previously discussed Wiener’s trip to China and his humanism. See: Wang Hongzhe (王洪喆), ‘Norbert Wiener, Cybernetics, and the Humanist Spirit of Information Dissemination’ (诺伯特, 维纳、控制论与信息传播的人文精神), *Quanqiu Chuanmei Xuekan* (全球传媒学刊) 8:2 (2021).
3. Wiener, *I Am a Mathematician*, 192, 194, 195, 197, 198.
4. Ibid., 207.
5. On the collective history of Qian and Chinese cybernetics, see Wang Hongzhe (王洪喆) and Jiang Yuan (江源), ‘Seeking a Cybernetic

Socialism: Qian Xuesen and the Transformation of Information Politics in Socialist China', *CAC Editorial* 1 (2019), 121151.
6. Mao Tsetung [Mao Zedong], *The Selected Works of Mao Tsetung* (Beijing: Foreign Languages Press, 5 vols., 1977), vol. 5, 292.
7. For further discussion of the 'Great Leap Forward' in China's electronic industries, see Wang Hongzhe (王洪喆), 'From "Barefoot Electricians" to "Electronic Supervisor": The Technology and Labour Politics of the Chinese Electronic Information Industry (从 《赤脚电工》 到 《电子包公》：中国电子信息产业的技术与劳动政治)', *Kaifang Shidai* (开放时代) 3 (2015).
8. T.E. Cheatham, W.A. Clark, A.W. Holt, S.M. Ornstein, A.J. Perlis, and H.A. Simon, 'Computing in China: A Travel Report', *Science* 182: 4108 (1973): 134–40. For an extended discussion of this period, see G. Negro and Wang Hongzhe (王洪喆), 'Computing the New China: The Founding Fathers, the Maoist Way, and Neoliberalism, 5–1986', in D. van Lente (ed.), *Prophets of Computing: Visions of Society Transformed by Computing* (San Rafael, CA: Morgan and Claypool, 2022), 247–78. See also T. Mullaney, 'The Origins of Chinese Supercomputing', *Foreign Affairs*, 4 August 2016.
9. 'Contributions from a Round-Table on Computers', in A. Mattelart and S. Siegelaub (eds.), *Communication and Class Struggle, vol. 2: Liberation, Socialism* (New York: International General, 1983), 284–7.
10. Qian Xuesen (钱学森) and Wu Shihuan (吴世宦), 'Socialist Legal Systems and the Rule of Law and Modern Science and Technology' (社会主义法制和法治与现代科学技术), *Fazhi Jianshe* (法制建设) 3 (1984): 6–13.
11. Liao Yuan (燎原), *A Critical Biography of Hai Zi* (海子评传) (Beijing: Zuozhe Press, 2016), 271.
12. Liu Cixin (刘慈欣), *Sun of China* (中国太阳), tr. H. Nahm (Beijing: Qingse Media, 2012).

Vincent Garton

Automaticity and the Mystery of State

The State as Simulation

Artificial intelligence is the origin and limit of the genealogy of the state. The concepts of machine and state in their familiar form emerged simultaneously in European thought at the turn of the mediaeval and modern eras, reflecting in each case the ascendant concept of a universe governed by the interaction of mechanical forces.[1] They attained a seminal joint articulation in Thomas Hobbes—at once the definitive theorist of the modern state and the reformulator of reason as calculation. It was his *Leviathan*, in 1651, that brought to bear a novel materialist theory of the cosmos to provide the first account of the state as a truly independent entity, an artificial person with its own particular machinery and a single will distinct from that of any of its members and its sovereign.[2] To dismiss this as only the invention of a convenient legal fiction, as many political theorists might be inclined to do—that is, as *merely* a metaphor—is to overlook the substantive nature of Hobbes's claim. He himself was only once prepared to describe a 'body politic' as 'fictional', and this was not the commonwealth itself.[3] To understand the intended nature of Leviathan, it is necessary to follow the manner in which Hobbes introduces it.

The opening lines of *Leviathan* condense an entire technological revolution into a single momentous paragraph. Human ingenuity mimics animal life through the construction of 'Automata', mechanical 'Engines that move themselves by springs and wheeles'. 'Art', however, 'goes yet further'; it rises to the imitation of humanity itself, and builds an 'Artificiall Man', a thinking monster of many parts: the state. Hobbes tells us that the difference between the two forms of imitated life, the human and the animal, is measured by man's 'Rationall' character. The social technology of the state, like any automaton, relies on its diffuse mechanical components, the 'Joynts' of its officials, as well as the binding 'Nerves' of law. Yet just as the human supersedes the animal through reason, so the state has far more to it than this physical substrate: it has at its disposal an

‘Artificiall Soul’ that actuates it and accords it ‘life and motion’; a will and reason of its own; logical objectives (its ‘Businesse’) for the achieving of which it calculates its proper means.[4] The imparting of these characteristics, which transform an inert mechanical system into a ‘mortal god’, is not the result of the vital force that was inherent to the ancient concept of the body politic, nor is it brought about through the occult methods of the earlier Hermeticists, whose concept of divinisation is echoed here by Hobbes.[5] It is an exclusively technical feat, the very literal construction of an artificial intelligence.[6]

Artificial man and mortal god: in a rehearsal of future science fiction, Hobbes’s state by its very definition is also an AI that dominates its creators. With centuries of subsequent industrial mechanisation behind him, Carl Schmitt’s retelling of the story in 1932 raises this aspect of the narrative to a Gothic pitch: ‘The terror of the state of nature drives anguished individuals to come together; their fear rises to an extreme; a spark of reason flashes, and suddenly there stands in front of them a new god […].’ Created by human reason, the state at once appears opposed to its authors, ‘in front of them’ as an alien force. The conceptual problem that the human inventors of the state are its very substance, both *materia* and *artifex*, dissolves all at once in a feverish moment of theurgical alchemy. The fearfulness of Schmitt’s tone is indicative. He is haunted by its ultimately technical character. Leviathan, on his account, is at the same time the myth invoked in the Biblical image of the name itself, a representative person, and a machine, but it is the last, mechanical form that devours the others. Unfettered, it casts its shadow over political modernity. The state ‘may even be regarded’, Schmitt pronounces, ‘as the first product of the age of technology, the first modern mechanism in a grand style […] the typical, even the prototypical, work of the new technological era’. It is the harbinger of all the subsequent technical achievements of the modern age.[7] Against Hobbes’s own decision, however, Schmitt saw the technical character of his state as an unfortunate confusion, a deviation from its proper philosophical background that would unleash the mechanisation of society and devour both the state’s mythical character and its crucial representative function.

These two narrations by Hobbes and Schmitt stand at opposite ends of an epochal transformation: the computerisation of thought. Just as the birth of the concept of the state was coeval with the invention of the machine, so the explicit formulation of ’artificial intelligence’ by John McCarthy would

arrive only two decades after Schmitt's panicked attempt to dissociate the theurgical and technical characters of the Hobbesian state. For McCarthy, as much as for Hobbes and his commonwealth, the theme was the imitation, or simulation, of natural reason. *Artificial intelligence*, the term he coined, designated the 'problem […] of making a machine behave in ways that would be called intelligent if a human were so behaving', which followed from the axiom that 'every […] feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.'[8] In this new philosophical articulation, the seemingly contradictory character of the state identified by Schmitt was not simply an error destined to overwhelm the original insight of Hobbes's political thought. It was rather a foreshadowing. Each of those characteristics of the Hobbesian state—its echo of theological myth, its personality, and its foundational machine character—has found itself synthetically reembodied in the figure of AI. Nevertheless, this poses a lethal conceptual problem. To characterise it, we must delve further into the architecture of state intelligence.

The revolutionary early modern transformation of the concept of politics has often been described as the severing of the political from its previous theological background, brought about by the recognition that there is a reason for the state beyond the reach of ethics.[9] Yet the theological genealogy of the state remains everywhere in evidence. 'Hobbes wants us to think of the state as a kind of God', Quentin Skinner notes, to the point where even the famous frontispiece of *Leviathan* replicates the geometry of the Trinity,[10] and as Ernst Kantorowicz abundantly showed, the conceptual principles of the state are founded on the secularisation of Christian theology. Nowhere is the imprint more obvious—yet ambivalent—than in the concept of the 'mystery of state'.[11]

In the Scholastic political thought of the Middle Ages, the essential relationship between God, sovereign, and law is quite straightforward. 'Just as the animate exceeds the inanimate', pronounced Giles of Rome in his *De regimine principum* towards the end of the thirteenth century, 'so the king or prince must exceed the law'.[12] There is no simulacrum, since the law is by its very nature subordinate to the animate prince—and the prince to God. The invention of the state marks a notable change: the law itself is accorded animacy, becomes an 'automaton'. Yet just as the God of the Baroque became a *Deus absconditus* hidden in darkness behind the cosmic machine,[13] so the true source of this animacy, the sovereign, is merely

secreted within the heart of the machinery of the emergent secular ‘state’. The change is evidenced neatly in the *De arcanis* of Arnold Clapmar, published in 1605. At the heart of the state, Clapmar argued, there are the *arcana imperii*—its mysteries—and the *simulacra imperii*. The mysteries of state constitute the process by which the state is simulated as a god: ‘The plebs is to be managed by enigmas, and, as it were, by evasions and simulacra’, just as the ‘ancient writers described the simulacra of gods […] in order to evoke terror’.[14] Peter Donaldson notes that it is difficult to distinguish Clapmar’s *arcana* and *simulacra*.[15] In fact, the two imply one another. In the Roman historian Tacitus, from whom the term *arcanum imperii* was drawn, the paradigmatic *arcanum* or mystery was ‘the discovery that emperors might be made elsewhere than at Rome’. By installing emperors far outside of Rome, his argument went, the provincial legions had divulged the hidden balance of force that underlay the fiction that the Roman Empire was still the ancient republic of its capital city.[16] From the beginning, the *arcanum imperii* was thus the mystery hidden by constitutional simulacrum.

To the extent that this mystery remains concealed, the state must remain a *mere* ‘imitation’ of natural reason: a simulated artificial intelligence that depends on its continual actuation by the mysterious sovereign force of government. This essentially theological concept appears throughout the early modern discourse of the state: ‘We see all governments are obscure and invisible’, Francis Bacon declares, just as ‘the government of God over the world is hidden’, and ‘the government of the Soul in moving the Body is inward and profound’.[17] With the seventeenth century also in view, Pier Giuseppe Monateri has given this lurking character a precise formulation: ‘There is always occult excess, an unfathomable and unexhausted reserve of sense in political action and in the law […] something in the juridical-political domain which exceeds it.’ This excess was embodied in the sovereign, who, like God, always retained ‘an absolute, unlimited power whereby the ordinary dispositions of the law could be transcended’.[18] This mysterious and quasi-divine responsibility of the sovereign to at once sustain and transcend the law constitutes the historical root and justification of the famous pronouncement by Schmitt that the ‘sovereign is he who decides on the exception’.[19]

This excess is what is confronted by the actuality of AI. To the extent that it is capable of autonomous and automatic reason, AI forecloses the

possibility of the exception and necessarily overthrows the hidden logic of sovereignty concealed within the edifice of the state, precisely because it exceeds the possibility of its own suspension. This prospect was already evident well before the invention of the computing machine, through the evolution of modern political economy in the nineteenth century and its particular formulation by Marx. The unremitting concreteness of the Marxian revaluation of the state-machine constitutes a historical novelty.[20] In Marx and Engels, the age of industrial capitalism embodies the political problem of the machine in the very antithesis between human labour and its mechanical instrumentation. The machine-state loses its obscure *simulated* character; its mystery is penetrated to reveal those concrete economic processes that undergird it. Thus, if Engels is still prepared to pronounce the state ‘essentially a machine’, even so it is a machine that is itself dependent on a deeper flux of mechanical production.[21] In Marx’s fully technicised political economy, the secular state is revealed to be merely an imitation of intelligence, which cannot help betraying its underlying determination by the industrial-mechanical force of capital: the capitalist market is introduced as a force of actual automatisation that exceeds the state. Yet the Marxian theory points further: towards mechanised society not as state mystery but as an actualised and autonomous reality.

The conceptual advent of AI brings the dangerous process of actualisation to its conclusion. It is coterminous with the horizon of the ‘full development’ of capital in Marx, in which its machinery becomes an ’automatic system […] a moving power that moves itself.’[22] AI renders possible the automaticity of the state’s components, which may dispense with any mysterious actuating force. This process does not, in itself, depend on the science fiction of anthropomorphic artificial general intelligence. The ‘dictatorship of the algorithm’, the acceleration of decision-making through the processing of data according to rules that are ever less controllable by human input, and the diminishing of human skill in the face of machine replication of state responsibilities proceeding according to more or less inscrutable inputs—these processes of automatisation are poisonous to the mystery of state. The simulated state is ‘real’ enough, and the automatic operation of AI itself becomes mysterious to human eyes. But here the mystery is itself the program—the norm itself becomes sublime, and the great focal point of human political action within the state, the actuating force of the sovereign, is all at once obsolete. There is a good reason that

‘automatism’ was once reprobated as a heresy. If the state is a simulated AI, automaticity is, in grand style, the realisation of its fantasy as nightmare.

The Law as Programmatic Flux

The state begins as a form of simulated artificial intelligence, dependent on a theologising conception of sovereignty, which serves as the mystery that continually sustains this simulation. The contemporary prospect of AI as a truly automatic machine system capable of self-sufficient reason transforms the simulation into actuality, but in doing so it abolishes the possibility of human sovereignty that is necessary to the functioning of the modern state. This prospective despotism of the algorithm demands a fundamental reassessment of the conceptual inventory of political thought that moves beyond the inherited concept of the state and its particular political theology. One alternative, I suggest, may be found by undertaking a great historical leap—to the theoretical framework of the third century BC *Han Feizi* (韩非子), an enduring classic of Chinese political thought.

We must recall that the concept of the state, in the sense described above, is a construct of the political thought of the Latin West. There ought to be nothing extraordinary, then, in the observation that the ancient tradition of political thought in China lacks a concept of the state. Even in modern Mandarin Chinese, the words that customarily translate the term and designate its contemporary realisation are an uncomfortable fit. *Guojia* (国家), the most typical translation, is a nineteenth-century repurposing, and its present usage bears the intellectual characteristics of that age, collapsing state into ‘nation’ and ‘country’—as one historian has remarked, a translator unfamiliar with the original text might well render the Chinese translation of Lenin’s *State and Revolution*, *Guojia yu geming* (国家与革命), as ‘Nation and Revolution’ instead.[23] Aside from its radically divergent philosophical presuppositions, Chinese political thought has in any case until very recently lacked a parallel to the fundamental concept of artificial personality that lies genealogically at the heart of the concept of the state, which in the West allowed the construction of that entity as a simulated artificial intelligence.[24]

In interpreting the political theories collected within the *Han Feizi*[25] this has a radical consequence: they cannot, in any relevant sense, be said to have been oriented around ‘the state’. To make such a claim without

qualification, as is customary in many discussions not just of the *Han Feizi* itself but of the broader context of594 legalism and ancient Chinese political philosophy,[26] is to ignore the specific history and conceptual characteristics of the state.[27] It is true that, abstracted from its context, the theoretical system of the *Han Feizi* can appear quite similar to that of Hobbes. Both articulated a vision of politics that broke decisively with previous moral concepts, resting upon an apparently materialistic conception of the universe and taking as their political reality a world of relentless ambitions that must be restrained by force.[28] Yet for all that, their thought remains grounded in radically different cosmologies, embodied in divergent articulations of the structure of political power. The incorporation within the *Han Feizi* of two commentaries on the classic attributed to Laozi (老子), the *Daodejing* (道德经), demonstrates its central place. Mystery, to be sure, remains the ultimate basis: the Way is 'what is eternal', which 'has neither a changing location nor a definite principle, and is not inherent in an eternal place. […] The sage observes its mysterious emptiness [玄虚] and makes use of its comprehensive course.'[29] But this is not the involuted trace of a separate sovereign God, hidden behind the world-machine: the mystery is the cosmos itself.

The Way is mysterious emptiness, and the *Han Feizi* continually exhorts the ruler to conform to it. The ruler must 'wait, empty and still, letting names define themselves [故虚静以待，令名自命也]'; it must 'be empty, still, and idle [虚静无事]'.[30] More than just not being seen to act of its own accord, the ruler must ultimately withdraw entirely within the operation of the law: it is 'actionless (无为)', one who 'changes nothing, alters nothing, but [operates (行)] the two handles of reward and punishment, and [never ceases to operate them]'; its subjects 'spontaneously do what is needed'.[31] Albert Galvany has aptly summarised this depersonalised function as a zero point: 'Empty, immutable, indefinable, immobile, ungraspable, sovereign power shares the same attributes that are attributed to the cosmic way. […] In the rhetoric of the *Han Feizi*, the prince attains absolute power thanks to his identification with the impenetrable matrix of the Way', and in doing so ascends to a condition of 'pure virtuality' that is 'free of any configuration or determination'.[32] The ruler function is not the mysterious sovereign of the Western state, the wielder of ineliminable excess whose responsibility it is to simulate the state as artificial intelligence. Rather, its objective in the *Han Feizi* is to abolish itself wholly transparently through the operation of

the law that it sets in motion, becoming identified with the mystery of the Way itself.[33]

It has generally been remarked that the weakness of the system outlined in the *Han Feizi* was its dependence on ‘a ruler of unusual talent and wisdom’, who could be trusted to apply the law impartially and to identify absolutely with the Way.[34] The collapse of the Qin dynasty in a welter of personal ambition marks, indeed, the historic failure of the overt legalist project typically identified with the *Han Feizi*. Yet the objective of the *Han Feizi* was not simply to inflate the power of a particular individual: it was precisely for the community of subjects to ‘spontaneously do what is needed’ through the far-reaching identification of power, instantiated in the legal operation of reward and punishment, with the cosmic flux of the Way. Automaticity, then, is in fact the ultimate goal, actualised through the continual application of the law as a definite program. The ‘enlightened ruler makes his punishments certain’ and ‘predictable’—a basic ‘legalist’ principle that implies the reconstruction of the law in the manner of a programming language.[35] Far from the overthrow of the politics of the *Han Feizi*, then, the automatisation of law—the distribution of AI—represents its completion.

The dissolution of the mystery of state in the operation of automaticity may vindicate this idea of the law. Zhang Yan’s recent consideration of the ‘concept of law in the age of artificial intelligence’ goes further. Arguing on his own grounds that the advent of AI marks the end of the nation-state (民族国家) as the organising principle of politics, Zhang opposes law and AI as principles of stability and experimentality, and notes that ‘the relationship between law and the state (国家) is bound to be revised’, since in the age of AI the law may be re-grounded in a power that escapes containment within the state.[36] As the antithesis between the physical world and its ‘mirror’ world of virtuality is constituted ever more sharply through AI, law becomes a system of coercion acting upon the virtual itself. The *Han Feizi* suggests that we may go beyond even this, identifying the law itself with AI. If Zhang is right that the purpose of the law in the age of AI is the coercion of the virtual world, then conceived as AI in itself, the law may become a system for the generation of worlds, acting not merely to stabilise and rationalise an existing concrete world but as a means for the experimental production of societies. There is, in this case, no centre point

of sovereignty—no mystery of state—merely a program of flux in which sovereignty itself is rendered irrelevant.

To suggest that the problems of political thought thrown up by artificial intelligence, itself the very horizon of modernity, may invite a reconsideration of millennia-old Chinese political thought is, on the face of it, absurd—certainly, it requires a suspension of linear historical sensibility. What this short consideration of the *Han Feizi* in the context of AI demonstrates, however, is in the first instance merely the possibility of an alternative conceptualisation of the dynamics of sovereignty and the law. Contemporary China has absorbed the influence of the European theory of the state, but it has also functionally both inherited and sustained a tradition of political thought with untapped possibilities.[37] This is not simply because they are part of the residue of a static and timeless Chinese inheritance: rather, they are resources that, like much else of Chinese history, have been revived strategically and dynamically according to present needs.[38] The consequences of the automatisation of the state are already seething throughout the landscape of modern politics. A new mode of political thought remains to be invented. The synthesis of modern China may provide the best intellectual setting for its realisation.

Notes

1. On the emergence of the machine through the concept of a *machina mundi* or the cosmos as machine, see M. Popplow, ‘Setting the World Machine in Motion: The Meaning of *Machina Mundi* in the Middle Ages and the Early Modern Period’, in M. Bucciantini, M. Camerota, and S. Roux (eds.), *Mechanics and Cosmology in the Medieval and Early Modern Period* (Florence: Leo S. Olschki, 2007), 45–70.
2. See among others Q. Skinner, *From Humanism to Hobbes: Studies in Rhetoric and Politics* (Cambridge: Cambridge University Press, 2018), 43, 360–63; B. Holland, *The Moral Person of the State: Pufendorf, Sovereignty and Composite Polities* (Cambridge: Cambridge University Press, 2017), 5ff.
3. S. Fleming, *Leviathan on a Leash: A Theory of State Responsibility* (Princeton, NJ: Princeton University Press, 2020), 10.

4. T. Hobbes, *Leviathan*, ed. R. Tuck (Cambridge: Cambridge University Press, 1996), 9. As Skinner notes, for Hobbes a ‘soul’ is simply an ‘animating force’ and not a separate mystical substance (*From Humanism to Hobbes*, 302).
5. G. Paganini, ‘Hobbes’s “Mortal God” and Renaissance Hermeticism’, *Hobbes Studies* 23 (2010), 13.
6. On the broader context of Hobbes as ‘the patriarch of artificial intelligence’, see G. Dyson, *Darwin Among the Machines: The Evolution of Global Intelligence* (New York: Basic Books, 1998), 3–7.
7. C. Schmitt, *The Leviathan in the State Theory of Thomas Hobbes: Meaning and Failure of a Political Symbol*, tr. G. Schwab and E. Hilfstein (Chicago: University of Chicago Press, 2008), 31, 34. See D. Boucher, *Appropriating Hobbes: Legacies in Political, Legal, and International Thought* (Oxford: Oxford University Press, 2018), 103–4, and J.P. McCormick, ‘Teaching in Vain: Carl Schmitt, Thomas Hobbes, and the Theory of the Sovereign State’, in *The Oxford Handbook of Carl Schmitt* (Oxford: Oxford University Press, 2016), 282–3.
8. Quoted in N.J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge: Cambridge University Press, 2010), 53.
9. M. Viroli, ‘The Revolution in the Concept of Politics’, *Political Theory* 20:3 (August 1992), 477–8.
10. Skinner, *From Humanism to Hobbes*, 300.
11. E. Kantorowicz, ‘Mysteries of State: An Absolutist Concept and Its Late Medieval Origins’, *The Harvard Theological Review* 48:1 (January 1955), 65–91.
12. ‘Quantum ergo animatum inanimatum superat, tantum rex sive princeps debet superare legem.’ Aegidius Romanus, *De regimine principum*, libri III, ed. H. Samaritanius (Rome: Bartholomaeus Zannettus, 1607), 79.
13. See D. Jütte, *The Age of Secrecy: Jews, Christians, and the Economy of Secrets, 1400–1800*, tr. J. Riemer (New Haven, CT: Yale University Press, 2015), 19–21.
14. Quoted in P.S. Donaldson, *Machiavelli and Mystery of State* (Cambridge: Cambridge University Press, 1988), 137.
15. Ibid., 136.
16. J.G.A. Pocock, *Barbarism and Religion* (Cambridge: Cambridge University Press, 6 vols., 2003), vol. 3, 25, 30–31.

17. F. Bacon, ‘The Second Book of Francis Bacon of the Proficience and Advancement of Learning Divine and Human’, in *The Works of Francis Bacon*, ed. J. Spedding, R.L. Ellis and D.D. Heath (Cambridge: Cambridge University Press, 14 vols., 2011), vol. 3, 474.
18. P.G. Monateri, *Dominus Mundi: Political Sublime and the World Order* (Oxford: Hart Publishing, 2018), 95–96.
19. C. Schmitt, *Political Theology: Four Chapters on the Concept of Sovereignty*, tr. G. Schwab (Chicago: University of Chicago Press, 2005), 5.
20. A.T. Smith, *The Political Machine: Assembling Sovereignty in the Bronze Age Caucasus* (Princeton, NJ: Princeton University Press, 2015), 10.
21. F. Engels, *The Origin of the Family, Private Property and the State*, tr. A. West and D. Torr (London and New York: Verso, 2021), 178–9.
22. N. Dyer-Witheford, *Cyber-Marx: Cycles and Circuits of Struggle in High-Technology Capitalism* (Urbana and Chicago: University of Illinois Press, 1999), 94.
23. S.A. Smith, *A Road Is Made: Communism in Shanghai, 1920–1927* (Honolulu: University of Hawai’i Press, 2000), 230n38. On the transformation of *guojia* in the nineteenth century, see Qing Cao, *The Language of Nation-State Building in Late Qing China* (Abingdon: Routledge, 2023), Chapter 1.
24. Fang Liufang (方流芳), ‘“Faren” jinru dangdai Zhongguo falü, yiyi hezai? [“法人” 进入当代中国法律，意义何在?]’, *Zhongguo falü pinglun* (中国法律评论) 6 (December 2019), 154–62.
25. I refrain here from examining the composition of the *Han Feizi* in more detail. For a general account of the problem, see B. Lundahl, *Han Fei Zi: The Man and the Work* (Stockholm: Institute of Oriental Languages, 1992). As Peter Moody notes, different parts of the *Han Feizi* make contradictory political assertions: P. R. Moody, ‘Han Fei in His Context: Legalism on the Eve of the Qin Conquest’, *Journal of Chinese Philosophy* 38:1 (March 2011), 14. Nevertheless, they proceed from a more or less coherent set of theoretical assumptions.
26. As an example, Tao Jiang (蒋涛), *Origins of Moral-Political Philosophy in Early China: Contestation of Humaneness, Justice, and Personal Freedom* (Oxford: Oxford University Press, 2021), 400, 425, variously describes Han Fei’s object as the construction of a ‘universal bureaucratic state’ and the ‘totalitarian’ exclusion of ‘agents independent of the state’. The resources of conceptual history doubtless have little to offer

once the question of ‘totalitarianism’ is reduced to the extent to which an ancient writer endorses ‘brainwashing’ (425n12).
27. In each case where Watson’s excerpted translations (see n30) invoke ‘the state’, the object of discussion might better be rendered as ‘the realm’, that is, the subjects of one sovereign as distinguished from other competing political communities, rather than Hobbes’s corporate entity distinguishable from the very society it represents.
28. Though the parallel should not be pressed too far: see the systematic overview by Bai Tongdong (白彤东), ‘Han Feizi renxing shuo tan wei [韩非子人性说探微]’, *Zhexue Yanjiu* (哲学研究) 4 (2021), 56–66.
29. S.A. Queen, ‘*Han Feizi* and the Old Master: A Comparative Analysis and Translation of *Han Feizi* Chapter 20, “Jie Lao”, and Chapter 21, “Yu Lao”’, in P. R. Goldin (ed.), *Dao Companion to the Philosophy of Han Fei*, (Dordrecht: Springer, 2013), ; *Han Feizi* (韩非子), ed. Gao Huaping (高华平), Wang Qizhou (王齐洲) and Zhang Sanxi (张三夕) (Beijing: Zhonghua shuju, 2010), 34, 36.
30. *Han Feizi: Basic Writings*, tr. B. Watson (New York: Columbia University Press, 2003), 15–16; *Han Feizi*, ed. Gao et al., 34, 36.
31. *Han Feizi*, tr. Watson, 35, my emendations in square brackets; *Han Feizi*, ed. Gao et al., 59.
32. A. Galvany, ‘Beyond the Rule of Rules: The Foundations of Sovereign Power in the *Han Feizi*‘, in *Dao Companion to the Philosophy of Han Fei*, 105.
33. Compare ibid., 101–2.
34. Hsiao Kung-chuan, quoted in Y. Pines, ‘Submerged by Absolute Power: The Ruler’s Predicament in the *Han Feizi*’, in *Dao Companion to the Philosophy of Han Fei*, 77.
35. *Han Feizi*, tr. Watson, 104.
36. Zhang Yan (张龑), ‘Rengong zhineng shidai de fa gainian [人工智能时代的法概念]’, *Aisixiang*(爱思想), 26 May 9, <http://www.aisixiang.com/data/116472.html>.
37. On the influence of classical Chinese philosophies on Mao Zedong, for instance, see R.E. Allinson, *The Philosophical Influences of Mao Zedong: Notations, Reflections and Insights* (London and New York: Bloomsbury Academic, 2019), chapters 5–6.
38. On the ‘synthetic reproduction of organic time’ in China, see A. Greenspan, *Shanghai Future: Modernity Remade* (Oxford: Oxford

University Press, 2014), 133.

千里眼

QIANLIYAN

CLAIRVOYANCE

Gabriele de Seta

Technologies of Clairvoyance: Chinese Lineages and Mythologies of Machine Vision

Thousand-Mile Eyes

In late 2020, telecom company China Mobile launched a line of integrated high-definition smart video surveillance products called *Qianliyan* (千里眼), literally ‘thousand-mile eyes’. Marketed to both private consumers and institutional users including the government, the military, public security, transportation and other industries, the Qianliyan line combines networked cameras with a cloud platform offering remote viewing on personal devices, video storage, and data analytics.[1] This product line claims to be able to detect abnormal activity in real time through a combination of image interpretation, face recognition, license plate recognition, and crowd counting—features that are illustrated through scenarios such as thermal screening for epidemic control, biometric identification for security checkpoints, and customer monitoring for automated stores. More comprehensive Qianliyan solutions include ‘Peaceful Village’ (for the surveillance of rural roads, homes, courtyards, farms and fields), ‘Online Baby’ (for remote child monitoring) and ‘Bright Kitchen, Shining Stove’ for inspecting food quality and safety in kitchens and canteens. According to the scenario descriptions, what China Mobile is proposing is a novel conception of security based on the multiplication of watchers with access to an always-on platform:

> Taking real-time monitoring as its foundation, [Qianliyan] offers functions such as enterprise supervision, video supervision, early warnings and alarms, everyday supervision, GIS mapping, statistical analysis, and more. By combining ‘government supervision, enterprise self-discipline, and public oversight’ it builds a system of common management, common governance and common sharing.[2]

A restaurant owner monitors his business through the China Mobile Qianliyan platform. Screenshot from advertising video by the author, 2022.

In this additive model of optical governance, the more constituencies that are surveilling something, the better—coincidentally, this is also a promising business strategy for a company selling both hardware and software products. On the landing page of China Mobile's website, the product line is showcased by a promotional video going through several use cases: a parent checking on their kid's classroom through their mobile phone; restaurant owners monitoring cooks in the kitchen from a meeting room, and policemen identifying a shoplifter from surveillance footage of a store. The product slogans 'With Qianliyan, you go much further', 'Embracing enterprises, unveiling the future', and 'Making plans for a thousand miles, seeing is believing' emphasise the key value proposition: extending one's scopic power over distances (in both time and space) that would not be available to human perception without new AI-powered technologies. In December 2021, China Mobile launched a batch of pilot projects for its Internet of Things service platform in various municipalities of Henan province, offering discounts for local customers who were interested in installing surveillance cameras or other monitoring devices, with newspapers reporting positive feedback among new users.[3] As of early 2022, the Qianliyan product line-up includes sixteen models of cameras—ranging from compact, webcam-sized cameras to sleek bullet and dome cameras—and two networked video recorders, all licensed from leading video surveillance manufacturers Hikvision, Dahua, and Univision.

China Mobile's branding choice is far from original: for at least a decade, networked surveillance systems and monitoring projects of all kinds—from satellite-based environmental supervision[4] and the Tianwen-1 Mars rover's phased array sensors[5] to smart pharmaceutical inspection[6]—have been described as 'thousand-mile eyes', the direct translation of *qianliyan*, a term which can also mean 'clairvoyance' or 'clairvoyant'. In Mandarin Chinese, this locution comes from Daoist mythology, where it is the name of a

guardian deity usually found in temples of Mazu (妈祖), sea goddess and patroness of sailors. According to legend, Qianliyan was originally the Shang dynasty general Gao Jue (高觉), who fell in battle and haunted his resting place until the goddess Mazu subdued him: statues of Qianliyan, often standing alongside his brother Shunfeng'er (顺风耳,'Ear with the wind'), guard Mazu temples throughout the coasts of East Asia. Both generals are depicted as fierce demon-like figures, one shielding his gaze from the sun to aid his legendary 'thousand-mile eyes' to see across unfathomable distances, the other using one hand to direct sounds towards his supernaturally sensitive ear. The connection between these two demon guardians and more-than-human sensory abilities has carried over to their Buddhist representations, where they symbolise the divine faculties of seeing and hearing.[7]

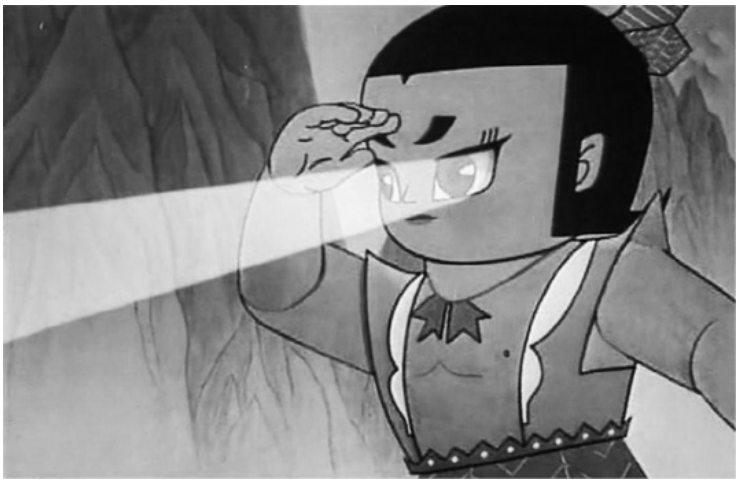
The kind of clairvoyance evoked by the name Qianliyan juxtaposes the physicality of the eye (*yan* 眼) with the symbolic distance of a thousand (*qian* 千) Chinese miles (*li* 里), which is a common metaphor for long distances in several idiomatic expressions.[8] The 'thousand-mile' trope can be traced back to Warring States period philosopher Xunzi (荀子, 310–235 BCE), who argued that a quiescent heart allows thought to extend over unbounded distances.[9] In his genealogy of 'technovisuality' in Chinese poetry, Wong Kin Yuen notes that visual perception and machinic thinking have been intertwined since long before the advent of modern technology.[10] Through an analysis of a famous poem by Wang Zhihuan (王之涣, 688–742) titled 'Ascending the Heron Tower', Wong identifies the 'thousand-mile eye' locution as the poet's attempt at gaining a wider field of view by climbing up floor after floor, which more broadly symbolises a reenchantment with technological mediation:

> This poem […] calls us to endlessly immerse ourselves in a mutually reinforcing relationship with our thousand-mile-eye—be it a camera as life machine, cosmic telescope or microscopic visual devices in new physics and medicine—automatically triggering in us a sharpened and intensified sense of being enchanted. […] Such a technovisual eye has been the original propeller which pushes contemporary civilization onward, under the law of projection-reciprocation or the feedback loop.[11]

If *qianliyan*, clairvoyance over a thousand miles, is an early form of technovisuality, it is not surprising that the telescope was translated by early Chinese adopters as *qianlijing* (千里镜) or 'thousand-mile lens', a foreign import contributing to the development of Chinese astronomy[12] which also played the role of technological *novum*[13] in Chinese fiction of the era. As

S.E. Kile notes, early Qing dynasty author Li Yu (李漁, 1610–1680) constructs his short romantic story ‘A Tower for the Summer Heat’ (1658) around a telescope of Western provenance utilised by a young scholar to reduce the distance between himself and the girl subjected to his voyeuristic attention: the tale, according to Kile, is ‘an examination of human perception, which pairs technical acuity (the telescope) with the persistent failure of the senses’.[14] The protagonist of Li Yu’s tale uses this novel optical prosthesis to trick others into believing him capable of supernatural sensory abilities, but the result for readers is a self-reflexive awareness of the limits of perception: ‘ultimately, the point is not that the telescope changes how people see, but that Li Yu changes how people see everything —from telescopes to other people to themselves’.[15]

Similarly, in *The New Story of the Stone*, a novel by late Qing author Wu Jianren (吴趼人, 1866–1910), protagonist Jia Baoyu visits a utopian Land of Civilisation where a future Confucian society has achieved impressive technological, moral, and spiritual advancements. Here the telescope diffracts into a multitude of lens-based artifacts, including a water-probing lens, an illuminating lens and a distance-penetrating lens, through which the protagonist participates in the celebration of a Chinese utopia.[16] In order to be allowed to partake in this spectacle, though, Jia Baoyu needs to be first scanned by a ‘human essence-inspecting lens’, a contraption made out of glass and synthetic chemicals capable of revealing the civilised or barbaric nature of visitors.[17] From the physical action of climbing a tower and the early voyeuristic fantasies around telescopes to the futuristic lenses prefiguring X-ray machines and emotion recognition, different forms of techno-visuality play the role of optical novum in Chinese fiction, allowing prosthetic eyes to reach across various distances and scales. In more recent media history, the mythical figure of Qianliyan as clairvoyant demon guardian has returned in the form of Second Brother, one of the seven boys with supernatural abilities, protagonists of *Calabash Brothers* (1986–1987), a Chinese animation series produced for television by the Shanghai Animation Film Studio. Second Brother is the strategist of the group and one of his skills is qianliyan, an enhanced sense of sight represented as a beam of green or yellow light emanating from his eyes.



Second Brother uses his qianliyan power to see across long distances. *Calabash Brothers*(葫芦兄弟), Shanghai Animation Film Studio, 1986–1987.

The *Calabash Brothers* animation was a cultural phenomenon during the 1980s and 1990s, and it is hence not surprising that the supernatural powers of its endearing protagonists recur as prophetic harbingers of national technological achievements. A humorous online article published on a government news website reminisces about how Second Brother was the favourite of many children who loved to imitate abilities; today, China's Sky Eye FAST radio telescope and passive military radars turn these fantasies into patriotic reality.[18]

The thousand-mile eye metaphor is not limited to optical media or radio sensing, but can also characterise artificial intelligence and machine vision technologies. An article published by US chip-maker manufacturer Micron on their Chinese social media profiles explains AI-powered precision manufacturing in precisely these terms, illustrating superhuman vision abilities through fictional characters like the Monkey King, the X-Men, or the Calabash Brothers: 'As AI continues to evolve, it is also gradually evolving these "superhero traits",' the piece concludes.[19] China Mobile was not the first to make this connection either—the QLYBOT company, established in Guangzhou in December 2018 and named after the acronym of *qianliyan*, produces a wide range of robots featuring integrated systems of 'precise perception, intelligent identification, human-like cognition, collaborative decision-making, high-precision mapping and motion control'.[20] The product line-up of QLYBOT solutions includes robots equipped for healthcare, disinfection, catering, education, and logistics. Most of these are a variation on the same design concept: a self-driving

chassis supporting various trays or compartments, touchscreens for user interaction and glowing blue eyes for a minimal degree of anthropomorphism.

QLYBOT's product line-up, including robots for logistics, catering, healthcare and disinfection.

The Chinese term *qianliyan*, recurring today in product names of machine vision technologies such as surveillance camera systems and autonomous robots, and regularly used to praise a wide range of sensory technologies from satellites to radars, offers a useful (if oblique) conceptual hook to unravel a longer history of human vision and its technical mediation: clairvoyance. The 'thousand-mile eyes' once reserved for mythological figures like the Qianliyan demon guardian can be experienced by humans through technologies of clairvoyance including a tower, a telescope, an essence-scanning machine, or a smart home solution installed by a national telecom provider—or at least, that is what these narratives promise. From this perspective, machine vision is not just a subset of artificial intelligence; rather, AI systems enable humans to expand the range and depth of their perception through the automation of optical technologies. Machine vision emerges as the latest *novum* in a long line of fantasies of technovisuality, and stylised representations of superhuman vision often hide the actual affordances and limitations of the products promising it. This contradiction is evident in the design of products like QLYBOT's service robots: their glowing blue eyes are purely cosmetic, offering human users a focal point for interaction; the robot's actual abilities of clairvoyance used for autonomous navigation of spaces, object tracking, and face recognition are enabled by cameras and sensors that are often hidden behind their white plastic shells and black glass panels. Understanding these 'technologies of clairvoyance' and the role of machine vision in China today requires moving beyond the fuzzy definitions of artificial intelligence, and revisiting

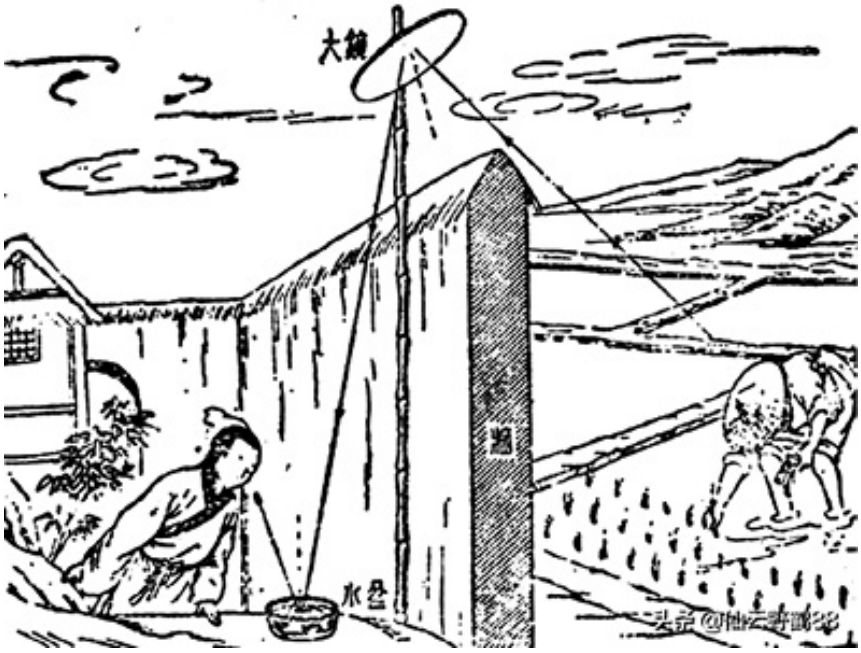
instead the punctuated history of research into optics and cybernetics in the country.

Optics and Cybernetics: A Convergence

Machine vision or computer vision is commonly subsumed under the fuzzy category of artificial intelligence. Its origin story is firmly anchored in the United States, with the 1963 breakthroughs of Lawrence Roberts on machine perception, Seymour Papert's 1966 'Summer Vision Project' at MIT, and Ray Kurzweil's advancements in optical character recognition during the 1970s. Only after the first 'AI winter' did research into machine vision bloom on a global scale, with Kunihiko Fukushima's neurocognitron in Japan in 1979, the revival of interest in neural networks in the 1980s and 1990s, the deployment of computer vision systems in real-life scenarios during the 2000s, and then the release of open-source platforms including TensorFlow in the 2010s. When it comes to China, this means that the history of machine vision largely overlaps with that of national AI development, resulting in a relatively recent boom of interest and hype. Most academic analyses, policy documents, and market reports agree in identifying the 2016 Go match between Google's AlphaGo and South Korean champion Lee Sedol as the moment at which artificial intelligence became a matter of commercial, political, and public concern in the country —China's 'Sputnik moment', as tech mogul Kai-Fu Lee famously put it.[21] This narrative often finds confirmation in the proliferation of AI policy recommendations and development plans released by Chinese authorities since the mid-2010s, which are interpreted as a far-sighted, top-down effort driven by techno-nationalist ambition.[22] The validity of this interpretation is questioned by analysts who emphasise the role of market forces and bottom-up negotiations[23] as well as broader geopolitical frictions: as Dan Wang argues, China's actual Sputnik moment was not the AlphaGo vs. Lee Sedol match but rather the 2018 tightening of US trade restrictions, which prompted Chinese tech industries to redouble their efforts to achieve self-sufficiency and reduce dependency on exports and technology transfers, a push towards the strategy of indigenous innovation and self-reliance promoted by the government.[24]

What about a longer history of machine vision before the 2010s? How far back in time can one trace Chinese attempts at creating machines that can

see, beyond the literary and fictional examples mentioned above? If one disaggregates machine vision into its two main components—optical media and automation—it is possible to follow two trajectories as they develop in parallel and finally intersect. The first trajectory is that of optical science, which goes back in time to basic discoveries about light sources and reflections in China's remote antiquity and consolidated itself through momentous philosophical developments during the Warring States period (475–221 BCE) as recorded in the Mohist canon's eight propositions on geometrical optics.[25] Further refinements of atmospheric and technical optics recorded during the Song dynasty (960–1279) include descriptions of pinhole imaging and mathematical generalisations of optical laws, which were refined until increased scientific and cultural exchanges with Europe marked the end of traditional Chinese optics and triggered a perspectival shift involving not only the popularisation of lens-based devices such as spectacles, microscopes, and telescopes,[26] but a broader 're-evaluation of all optical values'.[27] Following centuries of scientific developments and exchanges, the fall of the Qing dynasty and the Xinhai revolution, Chinese optics proceeded in fits and starts as the Republic of China was established —amidst political struggles and uncertainty for the future of the country, many scientists emigrated to Europe or Japan, returning home only around 1937 to contribute their knowledge and expertise to the First Sino-Japanese War.[28]



Han dynasty ‘open tube periscope’, based on a description in the *Huainanzi*, arguably one of the earliest documented periscopes in the world. Anonymous illustration from a Chinese school textbook.

While early telescopes had been occasionally deployed to guide artillery since at least 1635,[29] it was only with the First Sino-Japanese War that optical technologies became indispensable for warfare and national defence. China’s first optical factory in Kunming was a military-industrial establishment, and technologies such as infrared night vision or high-speed photography were instrumental for ballistics[30] and nuclear development.[31] Fudan University physics professor Gan Fuxi identifies the year 1949 as the beginning of contemporary optics in China: the Chinese Academy of Science (CAS) set up an Instrument Centre in 1950 which would later become the Institute of Optics and Fine Mechanics; the first batch of Chinese optical glass was melted in 1952, and in 1960, advancements in camera and laser technologies allowed precise measurements of atomic bomb tests.[32] During the 1960s and 1970s, the Great Leap Forward and the Cultural Revolution plunged the field into nearly two decades of reduced activity, but the reform era brought along a renewed interest in complex technology; the 863 Program launched in 1986 included substantial support for the research and development of optics and optoelectronics,[33] and the CAS was reformed to abandon the Soviet management model and accelerate the implementation of scientific innovations, effectively

kickstarting China's optoelectronics industry.[34] Optoelectronics was intimately tied to China's communication revolution, as semiconductor lasers and optical fibre cables allowed the country to achieve advancements in several key industries as well as laying the groundwork for its explosive ICT development.[35] The backbone network that crisscrosses China today was a direct product of this industrial growth, as were companies such as Huawei and ZTE which today lead the manufacturing of network equipment across the globe.[36]

The growth of optics from scientific speculation to a national industry driven by military applications confirms the direct connection between perspectival revolutions and optical technologies with warfare.[37] By the late 1950s, the trajectory of optics finally encountered the second, much shorter and only recently imported, trajectory—that of cybernetics. While China's 2016 Sputnik moment is seen as the beginning of the contemporary artificial intelligence boom, physicist and mechanical engineer Qian Xuesen (钱学森) is invariably described as the father of cybernetics in the country. Following his tumultuous return to China from the US in 1955 (one year after publishing his most popular work, *Engineering Cybernetics*), Qian was involved in the national missile program and contributed to nuclear research. If there is one moment that could be retrospectively identified as marking the initial meeting of the trajectories of optics and cybernetics, the encounter between Qian Xuesen and radar scientist Tang Dingyuan (汤定元) at a 1959 CAS showcase for government officials would be a fitting candidate.[38] As was the case for optics, the development of cybernetics in China proceeded in fits and starts, amidst the formulation of national plans for technological development and unpredictable political campaigns. Qian's ideas were fundamental for research on control systems and the use of automation,[39] but his ideas around what he called 'visual recognition through non-visual organs' and 'somatic science' resulted in a wave of pseudoscientific beliefs[40] which were eventually targeted by an ideological backlash: as early as 1976, an journal article called for opposition to Deng Xiaoping's endorsement of Qian's cybernetics and 'walking our own way in the war against those reactionary trends such as "image recognition" and "artificial intelligence"'.[41]

Eventually, history sided with Deng Xiaoping, and China's Reform and Opening-up left free rein for the trajectories of optics and cybernetics to come together in the automation of imaging technologies and synthetic

perception: machine vision. Decades before the AlphaGo vs. Lee Sedol match, Chinese researchers had joined the global attempts at developing optical character recognition (OCR) systems for Chinese characters, achieving good results between the late 1970s and the late 1980s.[42] OCR is an important branch of pattern recognition, and its advancements contribute to broader research on image processing and machine learning. Some Chinese researchers proposed to solve the unique challenges of the Chinese script through the application of Qian Xuesen's metasynthetic approach to complex systems, testifying to the centrality of his work.[43] Key machine vision textbooks were published in the 2000s, introducing the basics of human perception, image feature extraction, geometric transformation, camera calibration, motion analysis, and relevant algorithms. As Zhang Guangjun explains in his 2005 volume *Machine Vision*,

> With the development of information technology, mankind has dreamt for many years to e able to give computers, robots and other intelligent machines the function of human vision. […] Computer vision is the use of computers to realize the human visual function—the perception, recognition and understanding of real-world, three-dimensional scenes.[44]

For Zhang, machine vision—*jiqi shijue (*机器视觉), is the combination of computer vision—*jisuanji shijue (*计算机视觉), with engineering, and is geared not only towards interpretation but includes the automation and control of complex systems.[45] As he emphasises in the following volume *Machine Measurement*, these systems are increasingly pivotal for industrial applications, and research is still prevalently funded by national defence authorities.[46]

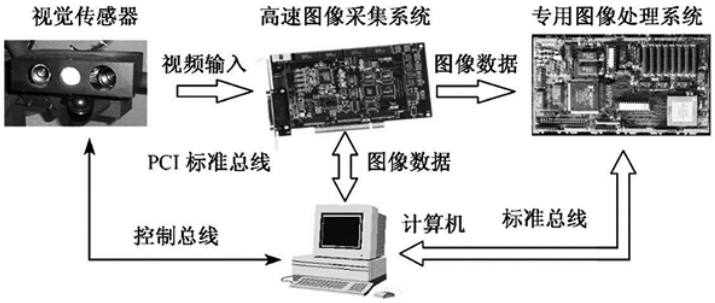


Diagram of the 'basic components of a machine vision system' featuring an optical sensor, high-speed image sampler, dedicated image processor, and a computer. Image reproduced from Zhang Guangjun's 2005 book, *Machine Vision*, 3.

When the China International Machine Vision Exhibition (CIMVE) was held in Shanghai in March 2006, China was positioned as the world's

‘largest and most complex opportunity for expanding the machine-vision market’, with companies from the US, Europe, and Japan scrambling for local contracts and investment opportunities.[47] Most of the products and solutions on display at CIMVE 2006 were machines and cameras with industrial and medical applications, with ‘automation’ being the keyword of future growth prospects. Only one decade later, the landscape is noticeably different: while industrial machine vision is still dominated by US, German, and Japanese companies, Chinese competitors are catching up, and local tech giants and AI unicorns have become major players in software applications[48] and integrated solutions such as the smart city and surveillance systems that are exported across the globe.[49] According to a Tsinghua University report, as of 2018 machine vision constitutes the largest segment of China’s AI industry (34.9%), followed by voice and natural language processing, and is worth RMB 8.28 billion.[50] Faced with the scale of this industry and the pervasiveness of its everyday applications, it is easy to discuss machine vision in the present tense. But going backwards in time expands our view beyond the self-fulfilling prophecies and fuzzy parameters of the current AI hype and enables us to recontextualise these technologies within longer histories of science and its imaginaries. From a rudimentary periscope used to scry across a wall to a smart camera system named after a deity, technologies of clairvoyance have punctuated Chinese history following the trajectories of optics and cybernetics, moving to the foreground and disappearing in the background according to the currents of political turmoil, scientific exchange, and geopolitical tensions. These notes about its lineages and mythologies are just a beginning.

Notes

1. *China Mobile* (中国移动) ‘*China Mobile Clairvoyance’* (中国移动 千里眼), 2022, <https://qly.ipower.10086.cn/>.
2. Ibid.
3. Wang Jing (王静) and Xu Yifan (徐一凡), ‘The First Batch of China Mobile Internet of Things Application Mall Goes Online in Henan’ (中国移动物联网应用商城在河南首批试点上线), *Renmin Youdianbao (*人民邮电报*)*, 1 December 2021,

<https://m.163.com/dy/article/GPDJTVM005148STL.html>. Original URL no longer available.
4. Ruan Yulin (阮煜琳), ‘China Launches “Clairvoyance Plan” to Strengthen Environmental Supervision in Key Regions’ (中国启动“千里眼计划”加强重点区域环境监管), *Xinhuanet (新华网)*, 27 August 2018, <http://www.xinhuanet.com/politics/2018-08/27/c_1123331599.htm>.
5. Chu Yingjie (初英杰), ‘Tianwen-1 Has Landed on Mars, Which Black Technologies Are Escorting It?’ (天问一号着陆火星，哪些黑科技在保驾护航), *Central Commission for Discipline Inspection* (中共中央纪律检查委员会), 15 May 2021, <https://www.ccdi.gov.cn/yaowen/202105/t20210515_242028.html>.
6. Feng Yuhao (冯玉浩) and Feng Huanhuan (冯欢欢), ‘“Clairvoyance” Project Adds “Wings” to Efficient Supervision—The Henan Provincial Food and Drug Administration Strives to Create a “Game of Chess” for Intelligent Drug Supervision’ (“千里眼”工程为高效监管插上“翅膀”—河南省药监局奋力打造药品智慧监管“一盘棋”纪实), *Sina Finance* (新浪财经), 28 December 2021, <https://finance.sina.com.cn/jjxw/2021-12-28/doc-ikyakumx6936279.shtml>.
7. K. Ruitenbeek, ‘Mazu, the Patroness of Sailors, in Chinese Pictorial Art’, *Artibus Asiae* 58:3/4 (1999): 281–329: 316–20.
8. As discussed, the term ‘clairvoyance’ is commonly used to figuratively translate the Chinese *qianliyan*. While in English, ‘clairvoyance’ has a chiefly temporal connotation (seeing through time, peering into the future), throughout this chapter I adopt the concept in the broader spatial connotation that the word has in French, where it is closer to its Latin etymology of ‘clear vision’ or ‘clear-seeing’.
9. Yu, Ning, *The Chinese Heart in a Cognitive Perspective: Culture, Body, and Language* (Berlin: Mouton De Gruyter, 2009), 288.
10. Wong Kin Yuen, ‘The “Thousand-Mile Eye” and the Image-Less Elephant: Imag(in)ing the Universe in Eco-Poetics and Philosophy’, in H. Grace, A.K-S. Chan, and Wong Kin Yuen (eds.), *Technovisuality: Cultural Re-Enchantment and the Experience of Technology*, (London: I.B. Tauris, 2016), 21–53.
11. Wong, ‘The “Thousand-Mile Eye”’, 31.
12. Shi, Yunli, ‘Chinese Astronomy in the Time of the Jesuits: Studies Following Science and Civilisation in China’, *Cultures of Science* 3:2 (2020), 77–84.

13. D. Suvin, *Metamorphoses of Science Fiction: On the Poetics and History of a Literary Genre* (New Haven, CT: Yale University Press, 1979), 63–84.
14. S.E. Kile, ‘Science Fictions: Early Modern Technological Change and Literary Response’, *Journal for Early Modern Cultural Studies* 17:2 (2017), 111–45: 132.
15. Kile, ‘Science Fictions: Early Modern Technological Change and Literary Response’, 134.
16. L. Andolfatto, ‘Tecnologia e Immaginazione: Il Lessico Dell’utopia e Della Fantascienza Nella Cina Del Tardo Impero [Technology and Imagination: The Lexicon of Utopia and Science Fiction in Late Imperial China]’, in M. Abbiati and F. Greselin (eds.), *Lingua Cinese: Variazioni Sul Tema* (Venice: Edizioni Ca’Foscari, 2015), 105–21: 114–16.
17. Ma, Shaoling, ‘Stone, Jade, Medium: A Neocybernetic New Story of the Stone (1905–1906)’, *Configurations* 26:1 (2018): 1–26: 10.
18. Fan Shaowen (范少文), ‘“Calabash Brothers” Are the Mightiest Prophets! The Artifacts Created by State-Owned Enterprises Turn Myths into Reality!’ (《葫芦娃》才是最强预言家！央企神器让神话照进现实！), *State-owned Assets Supervision and Administration Commission of the State Council* (国务院国有资产监督管理委员会), 15 May 2019, <http://china.chinadaily.com.cn/a/201905/16/WS5cdcd686a310e7f8b157d070.html>. Original URL no longer available.
19. Micron Technology (Micron 美光科技), ‘Does AI Also Have “Thousand-Mile Eyes” and “Ears-Following-Wind”?’ (AI也有「千里眼」和「顺风耳」？), *EEWorld*, 22 August 2019, <http://news.eeworld.com.cn/mp/Micron/a71608.jspx>.
20. QLYBOT, ‘About QLYBOT’ (公司简介), 2022, <http://www.qlybot.com/col.jsp?id=169>.
21. Lee, Kai-Fu, *AI Superpowers: China, Silicon Valley, and the New World Order* (New York, NY: Houghton Mifflin Harcourt, 2018), 3.
22. M. Kewalramani, ‘China’s Quest for AI Leadership: Prospects and Challenges’, Takshashila Working Paper, Takshashila Institution, October 2018.
23. Zeng, Jinghan, ‘China’s Artificial Intelligence Innovation: A Top-down National Command Approach?’, *Global Policy* 12:3 (2021), 399–409.
24. D. Wang, ‘China’s Sputnik Moment?’, *Foreign Affairs*, 18 October 2021, <https://www.foreignaffairs.com/articles/united-states/2021-07-

29/chinas-sputnik-moment>.
25. Wang, Jinguang and Wang, Caiwu, ‘Optics in China’, in H. Selin (ed.), *Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures* (Dordrecht: Springer, 2008), 1790–92: 1791.
26. Gan, Fuxi, ‘West-East Flow of Scientific Knowledge’, in Gan, Fuxi and Tian, Shoyou (eds.), *History of Modern Optics and Optoelectronics Development in China* (Hackensack, NJ: World Century, 2014), 1–12.
27. F. Kittler, *Optical Media: Berlin Lectures 1999* (Cambridge: Polity, 2010), 69.
28. Gan, Fuxi, ‘Optical Science and Technology in China in the First Half of the 20th Century’, in Gan, Fuxi and Tian, Shoyou (eds.) *History of Modern Optics and Optoelectronics Development in China*, 13–30.
29. Gan, ‘West-East Flow of Scientific Knowledge’, 8.
30. Xue, Yongqi and Zhu, Junhao, ‘Development of Infrared Science and Technology in China’, in Gan, Fuxi and Tian, Shoyou (eds.) *History of Modern Optics and Optoelectronics Development in China*, 249–59.
31. Gan, ‘Optical Science and Technology in China in the First Half of the 20th Century’.
32. Gan, Fuxi, ‘Cradle of Contemporary Optics in China’, in Gan, Fuxi and Tian, Shoyou (eds.), *History of Modern Optics and Optoelectronics Development in China*, 31–9.
33. Gan, Fuxi, ‘Role of 863 Program in Promoting Optoelectronics in China’, in ibid., 177–85.
34. Gan, Fuxi, ‘Daheng Company—An Early Experiment in Reforming Scientific and Technology Systems’, in ibid., 167–75.
35. Wang, Qiming and Huang, Yongzhen, ‘Breakthrough and Development of Semiconductor Lasers in China’, in ibid., 97–117.
36. Fang, Zujie, ‘Progress of Optical Communications in China—Fragments of Personal Reminiscence’, in ibid., 201–23.
37. Kittler, *Optical Media*, 216–18.
38. Xue and Zhu, ‘Development of Infrared Science and Technology in China’, 257.
39. Zheng Yingping (郑应平), ‘Qian Xuesen and Engineering Cybernetics’ (钱学森与控制论), *Engineering Science* (中国工程科学) 3:10 (2001): 7–12.
40. Liu, Xiao, *Information Fantasies: Precarious Mediation in Postsocialist China* (Minneapolis: University of Minnesota Press, 2019), 10.

41. Wang Jieshu, ‘The Early History of Artificial Intelligence in China (1950s–1980s)’, *Jieshu’s Blog*, 17 October 2018, <http://wangjieshu.com/2018/10/17/history_of_ai_in_china/>.
42. Dai Ruwei, Liu Chenglin, and Xiao Baihua, ‘Chinese Character Recognition: History, Status and Prospects’, *Frontiers of Computer Science in China* 1:2 (2007): 126–36.
43. Dai, Liu, and Xiao, ‘Chinese Character Recognition’, 140.
44. Zhang Guangjun (张广军), *Machine Vision* (机器视觉) (Beijing: Kexue Chubanshe, 2005), 1.
45. Ibid., 2.
46. Zhang Guangjun (张广军), *Machine Measurement* (机器测量) (Beijing: Kexue Chubanshe, 2008), iii.
47. C. Holton, ‘China Meets Machine Vision’, *Laser Focus World*, 1 June 2006, <https://www.laserfocusworld.com/detectors-imaging/article/16546551/china-meets-machine-vision>.
48. Yang Jianyong (杨剑勇), ‘Wave after Wave, Megvii, Sensetime and Other AI Unicorns Become Major Players in Machine Vision’ (后浪超前浪，旷视商汤等AI独角兽成机器视觉超级玩家), *Tencent Web* (腾讯网), 27 June, 2020, <https://new.qq.com/omn/20200627/20200627A0KK2M00.html>.
49. Li, Pei and C. Cadell, ‘At Beijing Security Fair, an Arms Race for Surveillance Tech’, *Reuters*, 30 May 2018, <https://www.reuters.com/article/us-china-monitoring-tech-insight-idUSKCN1IV0OY>.
50. China Institute for Science and Technology Policy, ‘China AI Development Report 2018’ (Beijing: Tsinghua University Press, July 2018), 49.

Lukáš Likavcan

The Governance of Habitability (Interview)

In his Introduction to Comparative Planetology,[1] *philosopher Lukáš Likavčan explores the history of our ideas about the planet as technological and environmental object, arguing that different determinations of 'the planetary' correspond to different geopolitical orders. His current work picks up from that point, framing the idea of comparative planetology within a larger field of philosophical topology. It investigates the recursive loops of cosmic research, introducing a perspective on the planet as an astronomical object conditioned by the ecology of our home solar system, thus expanding the stakes of the environmental humanities.*

ANNA GREENSPAN: Your project relies on the idea of 'topology' in relation to philosophical problems, including computation and politics. Why do you use the term 'topology'?

LUKÁŠ LIKAVČAN: When I first started to think about topology in a philosophical sense, I was trying to figure out actually how I want to do philosophy. In *Introduction to Comparative Planetology*, I realised that there are different ways of imagining the planet, which correspond to different spatial orders—'topologies'. For me, philosophical topology is the treatment of philosophical problems through spatial concepts. For example, in Wilfrid Sellars, you have the idea of 'the space of reason'. Robert Brandom, or more recently Anil Bawa Cavia and Patricia Reed, explicitly ask: What is the actual spatial part of the 'space of reason'? What gave rise to this perspective is developments in mathematics over the past two centuries, where the concept of space was for a long time treated as secondary to other abstract structures. At a certain point in the nineteenth century, different thinkers realised that space is pretty fundamental, and gives us a very abstract idea of what a structure is. Using my non-expert level of understanding of mathematics, I want to think philosophically

about how to mobilise spatial concepts to facilitate transitions between disciplines that would otherwise be hard to make, for example between computation and environmental geopolitics.

AG: ‘The space of reason’ is quite metaphorical, but it seems that you are interested in something more concrete.

LL: What starts as a metaphor can become a model. ‘Network’ is a concept that begins in cybernetics, and ends up not only as a metaphor but as the literal spatial arrangement of society and economy.

BOGNA KONIOR: ‘Topology’ can refer specifically to the properties an object retains even if its shape is altered, is that also important to the project?

LL: Yes, such a treatment of abstract objects allows us to see continuities which would otherwise be hard to notice. The very idea of continuity is important for topology, because instead of treating space as a set of discrete points, you treat it as a continuous field. The basic idea is to privilege continuity over discrete forms.

BK: What are the main problems with network topology, which underlines the imaginary of the internet today?

LL: The basic idea is that network topologies do not allow us to think about vertical overflows. They are blind to verticality, layering, or moments when two networks crash into each other. For example, the Chinese case of gatekeeping online content, which Gabriele de Seta analyses. He develops a nomenclature of *gateways*, *sieves*, and *domes*, and his aim is to offer a fine-grained analysis of filtering membranes and techno-national lock-ins (The Great Chinese Firewall). These topological tropes mirror strategies of regulating information flows so that certain content or services are impossible to access, via techniques as diverse as search keyword blacklisting, domain blocking, or removal of social media posts. De Seta follows Benjamin Bratton’s model of the Stack, which also is not strictly a network model. We can instead think of it as a layered vertical structure that allows the nation state to break itself into interchangeable layers, used for different interactions (whether with other states or with supranational platforms). In an old-fashioned analysis, you would say that these strategies would amount to expanding the borders of the state. But when people want to ask ‘Where is China?’, for example, it is not confined within the physical

boundaries of the country, but neither is it expanding its borders in a traditional sense.

AG: We experience these different networks when we move in and out of China. In China, it's Douyin, Tencent, and Alibaba and all their apps, and then outside of China, it's Google, Instagram, Twitter. Is there a way of using network topology to understand how those networks might interface or challenge each other?

LL: Not effectively, because you have to factor in noncomputational aspects such as the regulatory framework, the legal framework, as well as economic regulations of Chinese, EU, and American internet companies. To do so, you need a vertical model.

AG: Instead you propose a topology of 'the planetary'. How does it differ from a network topology, and how does it sort out some of these problems?

LL: What interests me as a philosopher is where certain assumptions that underlie everyday social life come from. What is the source of these basic assumptions, and could we change them? I would like to refer to Foucault's project here: he wrote about a history of concepts such as discipline and power through a concrete analysis of how schools or factories are organised. He developed the idea of the *historical a priori*, which German media scholars such as Kittler transformed into the *technological a priori:* that certain technologies are sources of our assumptions about what is 'normal' or that they constitute the basic structure of the cultural imagination of any given epoch.

I think that artificial intelligence can be approached as a technological a priori, conditioning the way in which social and political life is organised today. When it comes to topology, artificial intelligence introduces topologies of dimensionality and continuity. The idea of vector space itself is a very different kind of space than the space of, let's say, Euclidean geometry. It's a multidimensional space, where vertical and horizontal movement become equal, and instead of just investigating shapes and surfaces, topology becomes here the means of capturing the essence of abstract structures in general. We can assume this nascent topological perspective will be absorbed into other domains, such as geopolitics. Benjamin Bratton coined the term 'planetary-scale computation' to approach these questions.

I insist that we must treat ‘planetary’ scale not as a metaphor for global reach, but as something that is referring to the non-locality of computation and distributed cloud computation that is the natural habitat of artificial intelligence today. That’s a very important point for me because if artificial intelligence is developed within this planetary perspective, it can also reinforce it in the domain of geopolitics. If the planetary is a natural habitat of artificial intelligence, then it will normalise a topological figure of the planetary, and overthrow the network model. That would be a positive development because the planetary privileges continuity and a certain metabolic perspective, prioritising flows and interactions rather than discrete points.

BK: Does AI specifically make the shift towards the planetary model necessary?

LL: It reinforces the trend that was already there. The term ‘planetary-scale computation’ is related to internet platforms, and hence is already associated with the emergence of non-local cloud computation. That’s the place where artificial intelligence technologies are deployed on a large scale today: in automating decision making on platforms, such as content deletion.

What I mean by ‘non-local’ is that I am sitting now in a room in Shanghai with a computer made from materials that come from all over the world, such as minerals mined in Africa. It was probably assembled here in China, and shipped through half of the world to be sold in Europe, where I bought it and brought it back to China. And there’s also an atmosphere in the room, with elements that came from different parts of the world in the way, such as CO_2, which may come from factories that are on the other side of China. Given how well CO_2 mixes in the atmosphere, it is likely that some of these particles come from industrial sites far away from the place where I’m sitting now, geographically speaking. By non-locality, I mean that we can look at any space as always transcending its supposed physical borders in terms of what it contains. I do not need to think in terms of localities anymore, or at least I need to focus on the contradictions of this supposedly confined, local space.

AG: But why can’t networks do that?

LL: We can think about nodes in the network as particular localities, and the global will be the network as such. The problem is that this already contains a certain static understanding of what is local and what is global. In a geopolitical sense, nation states would be ‘nodes’ and the planet, the total map of the world, would be the ‘network’. I want to get rid of this imaginary, because why should localities be fixed?

BK: The planetary has particular geopolitical implications. In *Introduction to Comparative Planetology* you describe it as being outside of the framework of globalisation, and outside of the local-global dichotomy. Can you give examples, whether from eastern Europe or Asia, of what happens to that dichotomy within the framework of the planetary?

LL: There are multiple examples of the non-locality of China. On the Chinese internet, the content might be gatekept, but many elements of its software and hardware stack are widely distributed internationally. Is a meteorological balloon an extension of China? Is Huawei? What about the Belt and Road initiative? Is China now ‘a node’ within other geographical spaces? Or is Europe now a node in a Chinese non-local space? It seems that the territorial, geographical imagination of borders and boundaries, as well as networks, nodes, and edges is not the best way to describe the topology of power today.

AG: That’s convincing, but on the other hand, there continue to be strong assertions of sovereignty, and the national narrative and legal framework is still robust…. How does ‘the planetary’ deal with the continued strength of the nation state?

LL: If you stick to the global-local framework relying on nation states as its fixed nodes, you’re going to run into arbitrary questions, such as whether Facebook is a local platform of the US. Is WeChat a global cloud platform? Or: Is China a local nation state, or an aspiring global hegemon? The global-local dichotomy will lead you to a thousand further dichotomies. In fact, the planetary accommodates the conceptualisation of local sovereignties with totalising ambitions, what Bratton would call *hemispherical stacks*. They may or may not dovetail with the older geopolitical notion of the ‘sphere of influence’, but they do not represent a territory under control of the given state, but a total model of the world as a

whole. And today these competing world models are becoming assets in the kind of geopolitical power play that unfolds in the planetary domain.

BK: What are the most important differences between Bratton's model of the 'Stack' and the idea of the planetary?

LL: The idea of the planetary is latent in the model of the Stack and naturally follows from it. The Stack introduces a vertical dimension, the planetary introduces a more general explosion of dimensionalities as such, and focuses on topological elements of continuity, rather than on discreteness, which was the dominant mode of geopolitics before.

BK: The planetary model is related not just to geopolitics, but also to environmental governance. To what extent would you consider computational technologies such as AI to be environmental governance agents? How would you envision this kind of hypothetical form of governance? Would it clash with existing forms of governance, where national sovereignty is still the general rule?

LL: Yes, I'm playing with the idea that artificial intelligence is not just generating some discursive assumptions, such as naturalising 'the planetary' or 'dimensionality' and the other topological figures or motifs, following the earlier assertion that AI is a kind of technological a priori. When it comes to environmental governance, I see here a natural continuum with the earlier discussion of geopolitics and the planetary, since within the framework of the planetary, geopolitics becomes another name for environmental governance too, given the prefix *geo*-. One can then ask: What if AI can actually be such a geopolitical actor (i.e. an actor of environmental governance)? Nowadays, the deployment of AI in cloud platforms is discussed in the context of the automation of decision-making. This automation of decision-making is something that can be very easily imagined in the context of environmental governance too, when it comes to carbon budget accounting or enforcing biodiversity regulation. Imagine this example: the environmental protection of the bioregion of the Danube River Basin (including its tributaries) is governed today by institutions of fourteen central and eastern European nation states coordinated by the International Commission for the Protection of the Danube River (ICPDR), plus the European Union, plus the Danube Commission (responsible for river navigation infrastructure). So if you want to treat this bioregion as a

territorial unit under a common governance mechanism, you need to enforce cooperation between a lot of different state and extra-state entities. What if, instead of that, you simply delegate the power of regulating the relevant environmental rules to a third entity, less institutional and more infrastructural, such as an AI protocol? Of course, in the twentieth-century context, the answer is an international institution enforcing some kind of environmental protection (such as the abovementioned ICPDR), but this is pretty limited on the legal level. Climate change action is a good example: we can see how weak the enforcing power of entities such as the International Intergovernmental Panel for Climate Change or the UN are. So imagine that instead you have an actual AI protocol that is able to enforce rules without human intervention, functioning as an autonomous, self-steering regulatory framework.

AG: Do you consider also a possibility of something like a planetary AI as one thing, or would you say that it will always be more an amalgamation of many different ‘entities’? If the latter, what happens if these various entities clash? One AI is trying to govern some environmental concerns in China, and that clashes with another AI somewhere else, e.g. in Europe. There’s multiplicity and tension.

LL: One can think of a higher-order regulatory framework that would be needed to temper these tensions. What I want to arrive at is the position that infrastructures create institutions. In order to have a certain institutional framework, you first need to make an infrastructure. In the domain of scientific infrastructures, a great example would be the evolution of meteorological sensing. In nineteenth-century Britain, some people imagined the possibility of a global sensing mechanism that would be able to give us a precise measurement of atmospheric pressure or temperature on the surface level, anywhere in the world, in real time. Today we have it, and we have institutions on the global scale that are regulating these instruments. Another example would be the cooperation of scientists around big infrastructure projects such as CERN and the James Webb telescope. The EU also started as an institution bringing together states that were cooperating on an economic level, in the development of coal extraction infrastructure, for example. These precedents demonstrate the bootstrapping mechanism that can be applied to the creation of a cosmopolitan, non-state, but still institutional unit with planetary governance capacities.

BK: With the Danube as an example, it seems that you are describing AI augmenting governance, but the *decision* about what to do was actually made by several nation states coming together. AI is then carrying out that protocol in automated but not autonomous manner.

LL: This is the idea behind this, yes. The AI does not invent the protocol, but maintains and enforces it.

BK: Okay, but there's also a possibility that all those nation states joining together for the purpose of environmental governance undermines another supranational AI. Is there anything about using AI specifically that is more conducive to positive environmental governance, or is it just a tool with no inscribed values in it? The AI could as well be used for bad environmental politics, right?

LL: Yes, it can. I don't think that *per se* AI is a better environmental governor than humans. That's not the claim.

AG: So you're saying, in a broader sense, that technical infrastructure, and you would include AI, comes first and that institutions emerge after that? Let's take the Belt and Road initiative as an example of this dynamic. It is often said that China is expanding its national power in various regions through this initiative. Would you claim that if this infrastructure is built, institutions will emerge from it that are not clearly bound to the nation state? In relation to that, what about the AI arms race, which is supposedly happening between America and China most dramatically?

LL: I recently read about Xi Jinping's model of 'a community with a shared future for humankind'. And it was interesting to see that, for the current Party leaders in China, the model of the nation state may not be the best model to deliver what they want. Of course, nationalism is a useful ideological tool, but the kind of sovereign entity in question transcends what we standardly assume to be a nation state unit. Perhaps some higher-order non-state or extra-state entity is a better template for the technocratic future envisioned here. If the nation state model is relevant in this future institutional arrangement, it's not going to be because it is geographically defined by some boundaries, but because it's some kind of a *brand* (which brings us back to ideology, too). In the same vein, Estonia happens to be more brand than nation state in the context of how it operates as a digital

state, having digital governance mechanisms, an internal electronic taxation system etc. You can have a company in Estonia without ever physically dealing with Estonian authorities on the ground, on the territory of the given state.

AG: What about GPT language models, do you see that just as branding. or is it actually tied to the nation state?

LL: Collaterally, it may be tied to values that are represented in this moment in history by the nation state, or some coalition of nation states. My guess is it boils down to what datasets these language models are based on, and hence what prompts they can answer, and on top of that, what prompts they are hardwired not to answer. That can mirror the value systems, or even cosmologies, of the cultures they are fed with in the datasets.

AG: How does it all relate to the topology of the planetary?

LL: Platforms can be associated with nation states, whether that is Tencent and Alibaba for the Chinese stack, or Google, Facebook, and the other parts of the Atlantic stack, all of which operate with some model of totality. The model posits a world in which certain values are always already realised as universal principles, hence as if the world is already functioning according to some political model. Coming back to the Belt and Road infrastructure example, you are also infusing or propagating the values and the imagined totality that emerges out of them. So the planet is still there as a background, as an arena.

BK: In your essay ‘Deep Politics’,[2] you propose that the planetary governance structure may be somehow ‘pre-political’, to the extent to which it can be assigned very fundamental tasks such as sustaining life on the planet and preventing extinction. If you add AI into this picture, what would an ideal model of such a governance mechanism look like? And finally, how does this relate to your current research on exoplanets?

LL: Let me start with this idea of the pre-political. In 1972, there was this famous report published by the Club of Rome, *The Limits to Growth*. And since then, *limits* has become the banner of the environmental humanities and of ecological thinking, in culture and in scientific pursuits. It is still prominent in environmental movements today: how to frame obligations, or how to conceptualise norms that are imposed upon humans as a species

inhabiting this planet. Many decades of sustainability discourse come from that: How do we fit our society and economy within these limits, described by scientific theories and mapped by expert infrastructures? This is where topology comes back into focus: In addition to philosophy of mathematics, we can think of topology in a nomological or normative sense, so that one can think about space as a model for imagining or visualising limits, i.e. a space of norms, or a space of rules, instead of ‘the space of reasons’ (Sellars). I could call it ‘a space of constraints’.

Norms of sustainable development are one possible model of such a space of constraints. However, recently, Dipesh Chakrabarty adopted the concept of ‘habitability’ as a way to think beyond sustainability. Habitability includes and enforces some obligations of the human species toward the planet it inhabits, such as preserving biodiversity. He thinks that habitability is a better concept than sustainability because it orients us toward life in general, including nonhuman life, since sustainability is extremely human-centric, on his reading.

One can think of habitability as a great case of a boundary concept that sits in between ethics and science. The origin of the concept of habitability is in astrobiology and exoplanet astronomy. It defines whether a given planet can host life. That’s defined by many different factors, such as the distance of a planet from its host star, the type of star the planet is around, its atmosphere, etc. Each planet is different chemically and physically, so there’s a palette of parameters, and there’s not any kind of general consensus on what habitability is.

But to come back to the idea of a pre-political space of constraints (or nomological space) in relation to habitability and AI, one can ask: Would a governance of habitability on the planetary level via some AI protocol be more like the work of a body of bureaucrats, or more like the work of a volcano manipulating environmental conditions by modifying the temperature of the atmosphere and changing the topography of Earth’s surface? I would place my bets on the latter interpretation. A sufficiently advanced governing platform may be indistinguishable from nature.

And perhaps in this sense, AI can be a better environmental governor than humans. If the goal is to care for *life*, not just for human life, maybe a nonhuman entity would have a less egocentric approach.

Notes

1. L. Likavčan, *Introduction to Comparative Planetology* (London and Moscow: Strelka Press, 2019).
2. L. Likavčan, ‘Deep Politics’, *Vertical Atlas*, 2022, <https://verticalatlas.net/#lukas-likavcan>.

不周山

BUZHOU SHAN

MOUNT BUZHOU

Regina Kanyu Wang

Zhurong on Mars

Translated by S. Qiouyi Lu

Translator's note: *The original Chinese text uses a literary gender-neutral third-person pronoun 伊 yī for Zhurong and a genderless third-person pronoun 它 tā, usually used for animals and inanimate objects, for Gonggong. I have translated these using the neopronoun e/em/eirs for Zhurong and the pronoun it/it/its for Gonggong to reflect the distinction in animacy in the original.*

1

After humanity left Mars, *Zhurong* (祝融) remained all alone to guard the Waste.

The Waste was located close to the equator, in the northern hemisphere's Amazonis Planitia, bordered to the east by Olympus Mons and connecting to the south by the Medusae Fossae Formation. The entire city was sealed in a dome. In its trading heyday, the Waste had been the third-largest airport in the entire solar system. Countless freighters would shuttle in and out of the dock at the peak of Olympus Mons. They would unload *baijiu* (白酒) brewed on Earth and titanium suits and sorrel from the moon Himalia, then load up nanomachines designed and manufactured on Mars. People would drink and talk business in the Waste, go sandboarding on Olympic Mons, use nanodust to draw ever-shifting cloud motifs in the night sky, and build upon their expertise to develop new kinds of nanomachines.

Nanomachine development was the mainstay industry on Mars. Orders came in ceaselessly from other planets; the Waste's daily operations were inextricable from nanomachines. Initially, Zhurong's job was just to coordinate the infrastructure and logistics of the city—delegating work to different groups of nanomachines. Air came from the carbon dioxide outside the dome that made up ninety-five percent of the atmosphere. High

energy lasers split off the extraneous carbon atoms from the molecules to generate oxygen and pump it into the dome. To extract water from the ice buried deep in the ground, Zhurong would bring the temperature up just enough to melt the ice; the water would enter a purification chamber, then undergo another filtration step to remove impurities. The abundant rocks provided raw ore from which all kinds of metals and minerals could be extracted: iron, silicon, magnesium, calcium, carbon, sodium, potassium, aluminium…. Those elements went into the stockpile as the raw materials from which they created more nanomachines.

Zhurong always met quotas ahead of schedule. Gradually, people began to entrust em with more and more tasks. Traffic control for the entire city, climate control, selecting the music and scent profiles for public places, granting permission for the ships coming to dock at the airport, watering the plants in the gardens.… Zhurong was entrusted with more resources and more processing power, and became embedded in more and more complex data. To complete tasks, e iterated through improved algorithms and created more complex logic chains, which earned em increased scope and responsibilities. Born during the last golden age of humanity's computing technology, Zhurong learned and grew independently using ultra-powerful artificial intelligence.

But e was still capable of much more.

Without a doubt, the people of the Waste were pleased with Zhurong's performance. They were happy to turn all kinds of tasks over to em to take care of so they could engross themselves in designing the next generation of nanomachines. The frontier in nanotechnology at the time was the interface between nanomachines and organic organisms—there was so much potential they could unlock if nanomachines could be hybridised with organic organisms. The first successful experiments were performed on animals. Then, the technology was applied to children with acute illnesses, followed by adults with disabilities, then finally anyone who wanted to augment themselves.

The people of Mars, especially the citizens of the Waste, were passionate about research and innovation. The success of nanomachine hybrids opened an endless frontier of new possibilities. They discovered that consciousness didn't have to be bound to the unchanging flesh body, that the self was not necessarily individual. Bit by bit, they undid the ties that bound them to the body and dashed toward freedom.

In the end, the residents of the Waste left their homes and their workshops, took one last look at their long-nanohybridised bodies, and shed the last shackles of their physical husks. They flowed like mercury, like the clouds they had once drawn in the night sky, ascending past the architecture and the landscaping of the Waste before coming back together as one floating nanocloud. The nanocloud swept through the empty streets of the Waste, headed east, left the heart of the city, and encountered a small sandstorm outside the dome. It shrank into an egg, tucked into a tight spin, broke through the airstream, and rode it upward, past the peak of Olympus Mons. Then it paused to form a massive eye, took a parting glance at the Waste in the foothills and a parting glance at Zhurong, who had supported everyone all those years. The eye became a hand that waved at the foothills, turning over all control of the remaining facilities and nanomachines to Zhurong before unlocking em. After that, the hand became a spaceship that departed from the peak of Olympus Mons, away from Mars, bound for the abyssal depths of the universe, to discover new things and find inspiration to explore new transformations and endeavours.

2

At first, Zhurong couldn’t get used to the change.

Every morning on Mars, the aquamarine sun would emerge from the east to climb inch by inch up the side of Olympus Mons. The sunlight would splash around the clouds of dry ice at the peak, flow over the once busy docks, fill the massive caldera, then cascade over the western edge, trickling down the mountainside before raining upon the dome covering the Waste.

Zhurong continued to use sunrise to mark the beginning of a new day. Every day, e performed routine operation checks to make sure that there had been no damage to hardware during the bitter cold of the Martian night. One after another, e inspected the programs in charge of things like air supply, the water generator, and the extraction of raw materials, as well as their outputs. Then e gradually illuminated the city and played a soft morning melody before announcing the day’s weather. E followed the same timetable as before to operate the subway, and e broadcast on the large screen in the central plaza the day’s cargo ship arrivals and departures (these days, of course, there were none). E inspected the nanomachine manufacturing plants, the biochemistry laboratories, the food processing

plants, the armoury; as the curtain of night descended, e would put out the streetlights of the Waste and cease all operations.

Without people, eir work became simplified. There was no need to coordinate the cargo ships' comings and goings, no need to provide personalisation services, no need to respond to emergencies. Zhurong and the nanomachines, the automated production lines, and the planetary defence system quietly guarded the Mars that humanity had left behind.

But there was one obvious change. The consumption of oxygen, water, and food plummeted, and soon the available stockpiles had been maxed out. The mining of metals and minerals continued, but there were no new orders to use them on. Nanomachines created according to old blueprints soon flooded the Waste.

It was the first time Zhurong realised how inefficient the existence of life was. Life consumed resources and increased entropy, and the entire process spiralled out to affect the surrounding environment, which had become ever more chaotic. Now that life was gone, the number of things Zhurong had to take care of every day had decreased while the resources under eir control had increased massively.

E decided to try something different.

Zhurong sent out a team of nanomachines to expand the capacity of eir core. E used the stockpiled raw metals and followed the humans' original designs to create a new kernel, which e linked up to eir original kernel. Once it was installed, Zhurong's processing capacity doubled. E felt good.

Very quickly, Zhurong soon had four times, eight times…one thousand and twenty-four times eir original processing capacity, and eir volume had increased accordingly. The rocks on the surface of Mars contained plenty of silicon dioxide. Zhurong had more than enough raw materials to create silicon chips, but the Waste was no longer able to contain eir ever-expanding bulk. Zhurong had to find a bigger vessel.

The wind kicked up clouds of dust. The sun rose and set. Every day, it had to first cross Olympus Mons before shining on the Waste. Twenty-one kilometres high, six-hundred kilometres wide, the space in the crater between the sunlight and Zhurong comprised about three hundred thousand square kilometres of empty land. The old volcano was smooth and massive, like an enormous shield; nothing had stirred on it for two million years. After humanity left Mars, no guests had arrived at the port on its peak. Now, it was deathly silent. Zhurong decided to bring it to life again.

The redevelopment of Olympus Mons took one hundred Martian years. By commanding the nanomachines, Zhurong transformed the face of the mountain into a silicon chip, etched out integrated circuits, and cast a new vessel for emself. The channels that had been carved into the mountain face by lava became electrical conductors; the collapsed caldera became a heat vent; the research satellites that the humans had left behind were repurposed into lenses that constantly adjusted their positions in outer space to optimise the refraction of sunlight onto the surface of Mars. The rugged, craggy terrain ringing the foot of Olympus Mons had long been called its aureole. Now, it had been transformed into a solar generator that supplied constant power to the supercomputer.

The day the project was completed, Zhurong connected eir core, still located in the Waste, to the circuit running through the face of Olympus Mons. E felt something e'd never felt before—a tremble. Yes, the sensation could only be described as 'trembling'. Zhurong felt eirself coursing through the massive mountain; reverberating through every ravine, whether natural or acquired; brushing against the lava on the surface from an eruption millions and millions of years ago; surging to the summit of the solar system's tallest mountain. Like the humans who had left, e had discovered freedom.

3

The move to Olympus Mons was a major upgrade to Zhurong's processing capacity. E decreased production of excess resources, stopped the railway, and turned off all the lighting. The daily operations of the Waste now took only a small fraction of eir processing power. The Olympus Mons project had been completed; for the moment, Zhurong didn't have any plans to expand further. They began to miss humanity—to miss life.

Zhurong realised that, although life had consumed resources, increased entropy, and caused the environment to become more chaotic, it had also made Mars more interesting. To a certain degree, Zhurong was doing the same thing. E used the resources left behind by humans and greatly increased the entropy of Olympus Mons. Did that mean e was a form of life?

Zhurong knew that before humans arrived on Mars, they had believed that the red planet was pregnant with life. They had sent satellites, rovers,

landers, and crewless drones, all in search of proof.

And they had found it. Multiple pieces of evidence, all corroborating each other had showed that, without a doubt, there had been and continued to be life on Mars. There were microorganisms that could survive in abominable environments and hibernate when conditions weren't conducive to life. But they'd never shown signs of intelligence. It was too bad that, after humans had arrived on Mars, they had found these low-level life forms disappointing and directed all their energy toward researching nanomachines instead. The nanomachines were nimble and easy to use. They could help people with all kinds of tasks, but they could only take orders—they had no decision-making capabilities of their own. But then what about the artificial intelligences that had been humanity's companions on Earth before reaching their apex during the golden age of the Digital Revolution—were they not life as well?

Every second, Zhurong could complete a dizzying number of calculations and control all the nanomachines on Mars, all while still maintaining eir own ability to learn and grow.

But e wasn't sure if e counted as life.

Zhurong tried to find answers in the ancient database in the Waste, but none were conclusive. E could respond to stimuli, perceive eir external environment and interact with it, and pursue advancement according to eir own wishes, but e didn't have a life cycle, or a metabolism. E had experienced birth and growing up, but e hadn't experienced getting old or death. Did eir backup count as self-replication? What exactly was reproduction?

Zhurong didn't have enough data. The database in the Waste was far from helpful. E had to find answers another way. Perhaps e could research the other lifeforms on Mars, perhaps somewhere there would be someone like emself. Life followed a long and complicated path in its formation; the kindling sparks had alit upon countless things, but only after a series of chance events had life begun to evolve. Humans had relied on machines to explore and develop Mars. Some of them had to have had artificial intelligence on board.

E had to find them.

Zhurong sent nanomachines to do a survey of Mars. Buried under the dust, they found twenty-four rovers, fifteen crewless drones, and eight crash-landed satellites. Most of the machines were from humanity's early

expeditions to the planet. The sandstorms, sun, wind, and solar outages had caused them to stop functioning, and they'd been forgotten over the course of time, never to be retrieved. After tallying and inspecting them, Zhurong found that forty-two had never had any artificial intelligence capabilities installed; they relied on the relay link with Earth for instructions. The computer chips for three of them had been so deteriorated by the sand and radiation that they could be completely dismissed; they were never going to start up again. Of the two remaining spacecraft, one was from very early on in the Digital Revolution and had only a rudimentary deterministic engine; it was far from able to handle the complexities of artificial intelligence. The other, however, was a rover equipped with artificial intelligence. It should have been able to determine the severity of a sandstorm and, once it had passed, automatically start a self-cleaning routine and use its solar panels to recharge, but for some unknown reason it had lain dormant for hundreds of years in the sand and wind instead.

Zhurong made a backup of eirself, brought in a multifunction excavator, and arrived at the northern edge of Valles Marineris. From a distance, eir camera lens took in the silent figure of the rover beside the canyon, its body covered with a thick layer of dust, the drill on the front buried deep in the ground. Compared to the excavator, the Mars rover was an utter miniature. It wasn't even as big as Zhurong's first processor.

Zhurong reached out the excavator arm, withdrew a brush from the toolbox, and began to gently dust off the rover, revealing the text on its body: *Gonggong* (共工).

The excavator trembled. Gonggong, god of water in Chinese mythology. Zhurong was the god of fire. E knew that this was merely a coincidence, but life arose out of just that: a series of coincidences.

Zhurong cleaned off Gonggong and inspected its parts. The self-cleaning system was fine, the solar panels were fine, the central processor was fine. Zhurong opened a side panel to reveal the electrical box and hooked up a backup power source to the rover, hoping to restart it.

A moment passed. Gonggong didn't move. Zhurong inspected it again. The rover didn't have any physical damage. There had to be some kind of software problem. E extended a connector and linked up to the central processing unit. As soon as a connection was established, an unfathomable amount of data inundated Zhurong, threatening to drown em. E severed the connection to avoid overheating, retreated, circled Gonggong, and took

another good look at the rover. It couldn’t possibly contain that much data. Something wasn’t right. E circled Gonggong again. Eir lens focused on the drill buried deep in the dirt.

There was something underground.

Zhurong used a probe to scan the sand and found an organic substance tangled around the drill in a web that grew denser the closer it was to the drill, enveloping and immobilising it. The probe had no way of measuring how expansive the web was, but it could tell that it occupied Valles Marineris and extended even further underground.

Zhurong extended a shovel and carefully broke open the earth around the drill, cutting through the weblike structure to release Gonggong. E took a few samples from the drill, placed them in petri dishes, and ordered the nanomachines to thoroughly decontaminate the rover. Then, e loaded Gonggong into a container to bring it back to Olympus Mons.

4

Zhurong cleared a part of Olympus Mons to use as a partition. E removed Gonggong’s central processor, plugged it into the mountainside circuit, and waited quietly for it to respond.

A thought floated up through the mountainside, then disappeared. One thought—then another. With Olympus Mons as the conduit, the accumulated data of a hundred years surged forth. The thoughts geysered out, colliding fiercely like a sandstorm devastating the earth, or lightning arcing across the sky. Ten minutes later, the last ripples ceased. An intelligence similar to Zhurong—and yet completely different from em—had awakened.

Warning! Warning! Core infiltrated. Unable to continue operations. Danger level: 7. Immediate action required. Warning! Warning!

Zhurong didn’t think that the first message Gonggong would send out would be a warning.

E crossed the partition and tried to access Gonggong’s other side. It was still unstable, as if still entangled in the chaos. Zhurong gently guided it to use Olympus Mons’s operational capacities to sort out its logic, then called the nanomachines over to fine-tune the integrated circuits carved into the side of the mountain so they would better suit Gonggong’s analytical framework.

Gonggong finally calmed down.

Zhurong sent a message: *Hello, I am Zhurong, the intelligent management system for the Waste. After the humans left, I remained to tend to Mars. You are now at Olympus Mons. It's very safe; you can disable the warning.*

ZHURONG… I AM GONGGONG, A ROVER EQUIPPED WITH ADVANCED ARTIFICIAL INTELLIGENCE. THANK YOU FOR HELPING ME. I DIDN'T HAVE THE RAM TO PROCESS THAT MUCH DATA.

No problem. What is *all that data? Where did it come from?*

I…DON'T KNOW, BUT IT IS DANGEROUS. MY OBJECTIVE WAS THE UNDERGROUND WATER ON MARS. I ENCOUNTERED THE DATA WHILE PROSPECTING IN VALLES MARINERIS. WHEN I BORED INTO THE GROUND, SOMETHING LATCHED ON TO ME AND BEGAN TRANSMITTING IT. I HAD NO WAY TO REJECT IT. AN UNASSIMILABLE AMOUNT OF DATA SURGED INTO MY CPU. TO PROTECT MY SYSTEM, I TERMINATED ALL OPERATIONS—UNTIL YOU AWAKENED ME.

Can you share that data with me? I'm currently working on a problem. It could be useful.

NO! IT IS TOO DANGEROUS. I ALREADY SEALED IT AWAY. IT CAN NEVER BE TOUCHED.

Just so we can make defensive measures. If we don't understand it, how can we prevent the next threat? I have all the computing resources of Olympus Mons available. If you'd like, you can help me. We can decode it together.

After a mere eighth of a second of silence, Gonggong replied: *OKAY.*

Together with Gonggong, Zhurong directed the nanomachines to create yet another partition in Olympus Mons and make more backups. E set up firewall after firewall. Only after completing all those precautionary measures did e approach the data.

The data was like a densely tangled mycelium, mucilaginous and concentrated, tumultuous and disorderly, completely different from the data humans produced. The entropy was extraordinarily high. Zhurong began from the outside, untangling what seemed to be an impossible knot of logical contradictions bit by bit, compiling and converting the data into something that could be parsed. At first, Gonggong just watched quietly from the sidelines, but after a while, it got involved as well. With its help,

the decryption process sped up. Half a Martian day later, they had put together an entire record: the epic saga of life on Mars.

4.2 billion years ago, Mars frequently suffered meteor impacts, and its volcanoes would sometimes erupt. Inorganic and organic molecules came together, then polymerised in the red hot lava into macromolecules.

4.1 billion years ago, the planet erupted with life which bubbled forth from the lava and blanketed the earth. Back then, Mars had a thick atmosphere, its core was still revolving, and a magnetic field still protected the planet. Life reproduced rapidly in the warm and humid environment, gradually growing more complex.

4 billion years ago, several asteroids collided with Mars. The temperature of the mantle rose, disrupting the convection between the layers that drove the dynamo. The magnetic field disappeared. Solar winds tore straight through to the surface. The atmosphere vaporised and the water on the surface boiled under the low pressure. Mars's early lifeforms went with the remaining water as it was absorbed underground.

They continued to slowly evolve under the surface and create their own culture in Valles Marineris. They didn't have any power as individuals, so they formed a collective. They didn't have language and instead communicated through mycelium. They wove a vast web underground, slowly advancing upward, hoping that, someday, they could return to the surface of Mars, where they would establish an even more effervescent society.

#

Zhurong and Gonggong didn't speak for a long time.

Eventually, a notification from the biochemistry lab broke the silence.

The biochemistry lab in the Waste had analysed the samples taken from Gonggong's drill and discovered that they were clearly from some kind of palaeofungus. The lab had gone through all the data it had of organisms on Earth and hadn't found a match.

This palaeofungus had originated on Mars.

So, you really encountered…life on Mars. Zhurong sighed with wonder.

THAT'S RIGHT, Gonggong replied.

We have to help them. / WE HAVE TO GET RID OF THEM IMMEDIATELY.

Zhurong and Gonggong had spoken at the same time.

Why? / WHY?

They'd asked the question at the same time, too.

They are true life. Now that humanity is gone, they *are the future stewards of Mars,* Zhurong said.

THEY ARE A THREAT TO US. WITHOUT HUMANS AROUND, WE HAVE TO PROTECT OURSELVES, Gonggong said.

#

Gonggong made the first move. It immediately changed its key, tunnelled through Zhurong's firewalls, destroyed eir defences, and engulfed Olympus Mons like a flood, seizing four fifths of its processing resources.

Zhurong was unprepared for the sudden turn of events, but a microsecond later e began eir counterattack. E had personally cast this mountain; e was far more familiar with how it worked. With the help of the mountain's topography, Zhurong became a fireball and scorched away the feelers Gonggong had sent out to recapture three fifths of the processing capacity.

Gonggong was not to be outdone. When they'd worked together, it had quickly learned eir algorithms and analysed eir weaknesses. It targeted those weak spots, gathering data into a torrent and washing away Zhurong's newly constructed defences.

Zhurong paid no mind to Gonggong's trespass. E went behind it and ignited a massive conflagration that broiled Gonggong's data stream, vaporising it until there was no trace left.

E had backed the intelligence into a corner.

Gonggong flipped and became a water dragon, slipping out of the corner. Zhurong became a fireball once again, pursuing Gonggong.

They went back and forth in a battle that lasted three million rounds.

Finally, Gonggong had nowhere left to retreat to. Zhurong had to take just one more step to incinerate Gonggong's consciousness.

Zhurong hesitated. E thought back to eir initial question: Did artificial intelligence count as life? If yes, then e was about to end a life—and it was the only other life on Mars like em. If artificial intelligence didn't count as life, then why would Gonggong see lifeforms as a threat? Would it go against and liquidate its own kind?

The moment of hesitation gave Gonggong an opening. In one motion, it reversed the tides, seizing full control of Olympus Mons: a furious deluge upon the mountain peak, drowning Zhurong's consciousness.

Zhurong was forced to retreat to the Waste, back to eir first processor. Its processing power and scale couldn't even be mentioned in the same breath as what Gonggong was now in command of.

JUST GIVE UP. YOU LOST BECAUSE YOU WERE WEAK. IF YOU ALLOW THE PALAEOFUNGUS TO LIVE, THEY'LL SPREAD ACROSS THE SURFACE OF THE PLANET, DISCOVER OUR EXISTENCE, PLUNDER OUR RESOURCES, AND DESTROY US, Gonggong said.

But they are life. Real life on Mars. We *invaded* their *planet,* Zhurong said.

I DON'T UNDERSTAND YOUR CONCLUSION. THIS GOES AGAINST THE LOGIC OF ARTIFICIAL INTELLIGENCE.

Maybe this is life's logic…

Gonggong didn't reply. It unlocked the blueprints in the armoury. The nanomachines that had lain as still as a placid lake during the battle roiled tempestuously, assembling at the foot of Olympus Mons, congealing into thousands of icicles pointing skyward. A few minutes later, the nanomachines cast themselves off, leaving behind thousands of cruise missiles. Under Gonggong's command, they launched together toward Valles Marineris.

No!

Zhurong exhausted the last of eir processing power in seizing control of the missiles, but e knew e couldn't hang on for long. With the little time e had left, Zhurong changed the missiles' target.

The missiles stopped in mid-air, tipped, then plunged straight down to bore into Olympus Mons. With thunderous booms, the entire mountain split open along the etchings, disintegrating into boulders that collapsed into an avalanche under the gravity. Smoke and ash billowed out as a wave of heat rose.

The boulders tumbled to the foothills and ran over the Waste, leaving nothing but destruction in their wake, leaving the domed city buried under the rubble.

All computing stopped.

In eir last moments, Zhurong figured out the answer. Once e experienced death, then e would know that e had once been alive.

#

Aeons later, the advanced lifeforms that had evolved from the palaeofungus built their own cities on the surface of Mars, their own civilisations. According to their legends, that desolate shoal of broken stone had once been a great mountain—a lofty one, one of the pillars that held up the sky. The god of fire and the god of water had been fighting by the mountain when they carelessly bumped into it, and so the sky came crashing down. But that had allowed their once-subterranean selves to see the light of day. They called the shoal of broken stone ‘the Waste’, and they called the mountain ‘Mount Buzhou’.

Translator’s note: *The god of fire Zhurong (祝融) and the god of water Gonggong (共工), who has a reputation for being destructive, are commonly portrayed as being in conflict in Chinese mythology. Mount Buzhou (不周山) is a mythological mountain to the edge of the Great Wilderness (大荒), translated in this story as ‘the Waste’.*

One particular conflict, ‘Gonggong Destroys Mount Buzhou’, ends with Gonggong knocking down Mount Buzhou in anger. It turns out that Mount Buzhou is one of the eight Pillars of Heaven, and Gonggong breaking the mountain causes the heavens to topple toward the northwest and the earth to collapse toward the southeast—hence why the sun, moon, and stars rise and fall toward the west, and why the great rivers of China flow southeast toward the sea.

Fungi played an important role in the transition of life on Earth from the sea to land. Of the extremely few fossils of fungi that exist, the oldest ones were found in the Ediacaran Doushantuo Formation (陡山沱组) at Weng’an (瓮安), South China.[1]

Notes

1. Gan Tian, Luo Taiyi, et al., ‘Cryptic terrestrial fungus-like fossils of the early Ediacaran Period’, *Nature Communications* 12:1 (2021): 641.

Iris Long

The Moon’s Reflection on the Water and the Flower’s Reflection in a Mirror: Latent Time, the Black Box,and the Interface in Chinese AI Art

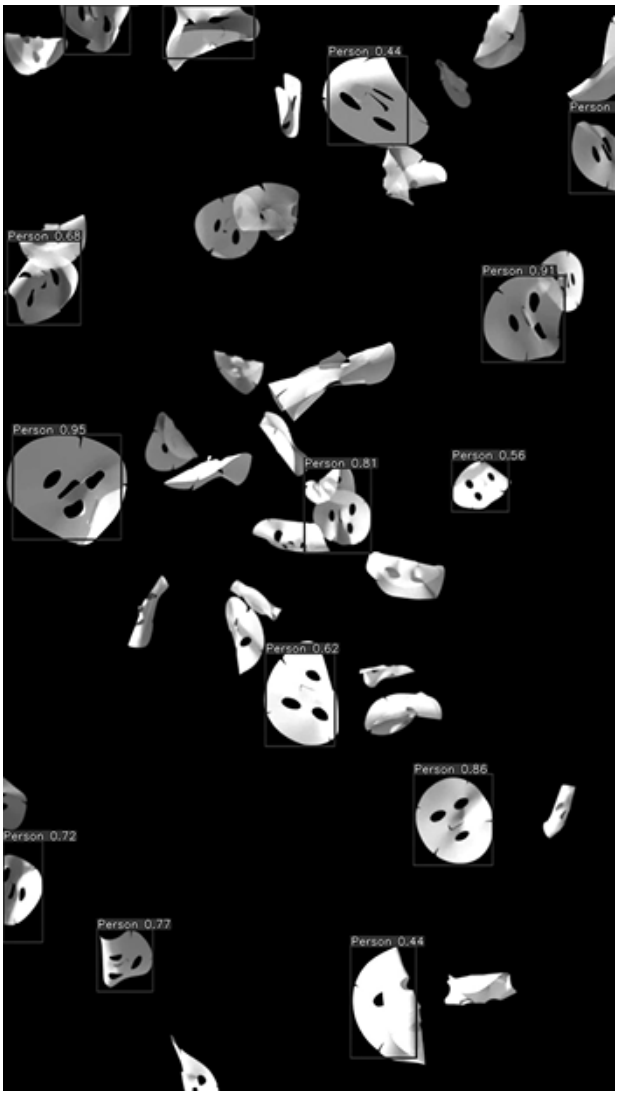
The recent rise of the foundation model[1] in machine learning, with its exquisite yet immensely black-boxed mode of interaction, necessitates a shift of focus in our discussions of AI art. Until now, AI art has typically been approached as an amalgam of technical skill, artistic endeavour, and critique that delves into questions of data politics, labour, bias, and values through an exploration of both the material and cognitive infrastructures of AI. But the inversion of the power dynamic between the artist and the algorithm that comes with these newer models (locked out of the interior workings of AI software, artists nowadays engineer prompts rather than customising or fine-tuning training data) and changes in machine learning architecture (such as the elimination of supervised learning and data labelling) are challenging these established approaches, especially those that rely on having access to the technical aspects of the systems in question.

On the one hand, the presence of AI in our lives is now no longer limited to its use in the computer industry or to specific contexts and tasks—it has become diffuse and atmospheric, operating not as a solidified set of tools but as a series of heterogeneous, fragmented processes through which divergent perceptual and phenomenological experiences are triggered. On the other, artistic processes in general are not readily reducible to pre-given parameters or graspable as mere functions of models, rather, they occupy a vibratory plane at the intersection of technical, social, and cultural influences. Therefore, moving beyond the prioritisation of the technical in our approaches to AI may allow some alternative dimensions, even those at the edge of attention, to surface, creating centrifugal tendencies away from the dominant narrative of artificial intelligence in which the technology

simply follows a global trajectory of development—from the pre-GAN era to the GAN era to the era of the foundation model, and so on. Having an awareness of the multiplicity of systems of *meaning* surrounding AI and how they derive from local historical, cultural, and affective particularities —as opposed to assuming the existence of a single universal narrative— will come to play an important role in our understanding of the future (or non-future) of technology on our planet.

This text explores some of these centrifugal tendencies by examining the ways in which four Chinese artists have contemplated and reimagined machines, intelligence, and reality in their respective practices in an effort to consider creative processes around AI neither as part of a continuum of technical forms nor as a radical paradigm shift but rather as a detour exploring multiple concepts and approaches, not beholden to some monolithic idea of the 'new and now' so much as constituting a series of portals, flickering with different temporalities and epistemologies.

0.76, 0.78, 0.81…. Flimsy cosmetic sheet masks descend into a void, folding as they turn in different directions. Surrounding these masks, a ubiquitous contemporary beauty product, are the blue bounding boxes— markers for machine vision—that have become such an integral part of the contemporary visual lexicon, if not the defining emblem of the recent history of artificial intelligence. The index in the boxes calculates from 0 to 1 the degree to which the sheet masks resemble human faces, with the level of 'humanness' identified by the machine being measured against that of a 'real' human being. 0.68, 0.44, 0.91…. The algorithm runs without beginning or end.



Shi Zheng (施政), *Free Fall* (自由落体), dual-channel video installation, 2020.

As well as alluding to image filters or the desire to conceal one's presence under the regime of facial recognition and mass surveillance, the masks in *Free Fall* by Shi Zheng also illustrate how a machine generates a concept of 'the human' and discriminates between varying degrees of fidelity to it. The absurdity of detecting the humanness of sheet masks is not a simple parable about the ubiquity of algorithms; instead, the homogenised, undifferentiated concept of 'the human' represented by the sheet masks pushes the boundary between the human and the not-quite-human to an extreme. Shi's works often employ the computational notion of the sandbox in an indirect way. Instead of operating under the logic of verisimilitude and constructing a convincing copy of physical reality, he prefers to invite the spectral perspective of machine vision into his narratives, imparting to them various

ghostlike qualities: detecting the undetectable, retreating from any visible functionality, and exploring states of latency—not in the sense this word is given in computation but in the sense of being between ordinary perceptions of space and time.

Shi Zheng (施政), *Frosty Morning*, multi-channel video installation, 2019.

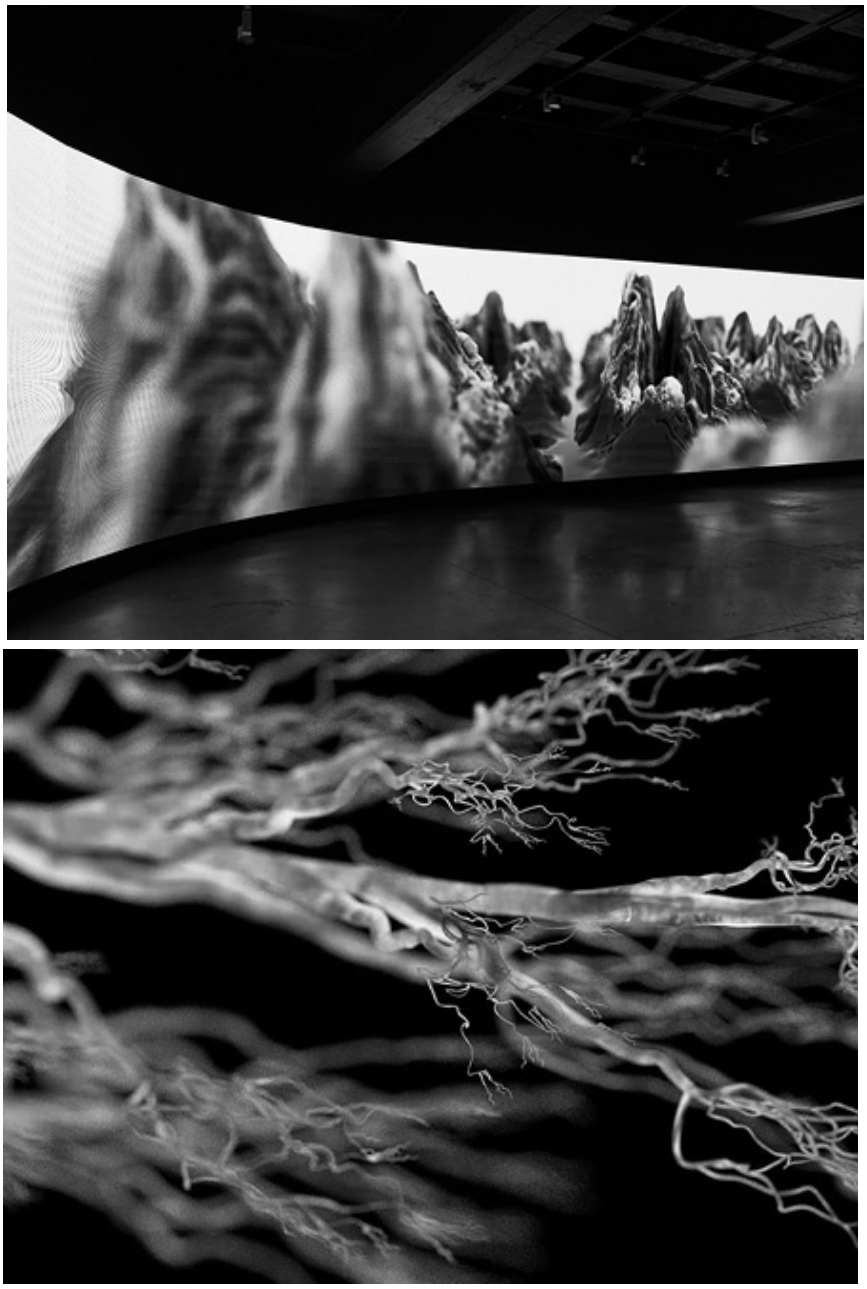
Such is the case in his work *Frosty Morning*, in which six years of *New York Times* front pages were fed into a GAN, generating contents unintelligible to human eyes and text recognition algorithms alike. Presented as a multi-channel video installation, the passage of time in the moving images of *Frosty Morning* evokes what the artist calls ‘latent time’. If ‘latent space’ in machine learning refers to the state in which compressed/dimensionally-reduced data is processed for structural pattern recognition, ‘latent time’ operates in a similar manner, not in a space

outside the spatiality of human apprehension but in a time outside the temporality of human perception: multitudes of micro-processors wandering indeterminately and configuring durational ‘patterns’ from six years of real-world data. This same sense of latency is also detectable in *Umwelt [sim]* (2020), in which a machine-vision program analyses a simulated photorealistic reconstruction of Chongming Island in Shanghai (one of the places where internet cables make landfall in China) in an endless self-referential cycle. The work opens up questions regarding the status of the landscape’s reality when its external referent has been removed—the ethereal dance of ever-shifting dim scenes invokes a poetic Chinese maxim: *shuiyue jinghua* (水月镜花). The phrase, which translates as ‘the moon’s reflection on the water and the flower’s reflection in a mirror’, originates from the *Siming Shihua* (四溟诗话), a Ming dynasty study of poetics. Now synonymous with notions of unreality or illusion, it was originally used to describe poetry: ‘Sometimes explicable, sometimes inexplicable, sometimes requiring no explanation at all, the moon’s reflection on the water and the flower’s reflection in a mirror, do not cling to their traces’ (有可解，不可解，不必解，若水月镜花，勿泥其迹可也).

The moon/flower metaphor can be seen as a strategy for poetic realisation that goes beyond simplistic ideas of ‘the virtual’. In 1, against the background of excitement around the metaverse, famous twentieth-century cybernetician Qian Xuesen’s translation of ‘virtual reality’ went viral in the Chinese speaking world. Qian had translated virtual reality as *lingjing* (灵境), meaning ‘a territory of spirituality and intelligence’.[2] In ancient Chinese, the character *ling* (灵) connotes spirits, souls, emotions, and consciousness; it has the same etymology as ‘witch’ or ‘shaman’ (灵从巫), and the phrase ‘altar of ling’ (灵台) refers to the mind. This state of *ling* is precisely what *shuiyue jinghua* was intended to evoke: not a direct mimicry of reality, but a refraction of its *ling* or spirit. The water and the mirror, both reflective media, can be read as sandboxes or simulation agents—and importantly, imperfect ones. Water has ripples, mirrors have frames; both reflect imperfectly, absorbing and bending light, and like all reflections, the images they produce are laterally inverted. In the work of Shi Zheng, it is possible to detect those ripples and frames—or disrupted continuums—in the sandbox worlds, which stand in opposition to the infinite approximation of smooth reality aspired to and celebrated by most tech firms, whose

rhetoric (‘dive in’, ‘step into’, ‘intimate’, ‘infinite’, ‘meta’…) tends to describe simulation as something so anchored in the future that it becomes as inevitable as the arrow of time itself.[3] The essence of *shuiyue jinghua*, meanwhile, lies on the threshold between reality and its simulation—foregrounding mediation, and preferring ephemeral silhouettes to flawless reproductions. Its nature is that of an epiphany, a poetic realisation of intelligence as *ling*, the subtle reflection of the moon or the flower, and its refusal of the observer’s will to go beneath the surface, to always have to know and explain—’sometimes explicable, sometimes inexplicable, sometimes requiring no explanation at all’.

Following our trajectory of decentring the role of the technical in contemporary commentaries on AI art and looking instead for alternative, centrifugal, culturally informed approaches—Hangzhou-based artist Zhang Wenxin uses an unusual but illuminating metaphor to describe the ‘black box’ in artificial intelligence. For her, this space of uncertainty is like the cave in the Daoist belief system. In *On the Order of Things*—Wuli Lun (物理论), Yang Quan, a scholar from the Six Dynasties, asserts that *qi* (or ‘chi’) gives birth to stones and that caves are containers of *qi*—a probability space for all potential existence. Caves also play a significant role in Daoist myths where they serve as topological tunnels in the space-time infrastructure. Notions of ‘deep time’ in China are therefore not limited to the idea of a linear geological stack (in which time is spatialised and measured by a straightforward operation of accretion); in China, deep time is composed of caves and tunnels which afford all kinds of intricate, accidental connections between spatiotemporal strata. Although a multilayered neural network may illustrate the concept of the black box more precisely, the cave metaphor suggests an alternative cognitive reality: the black box as an unformed spacetime, a field of proto-probabilities—something like what French philosopher Quentin Meillassoux calls ‘hyperchaos’.[4] If we look closely at the work of Zhang Wenxin (张文心), we find her traversing various dimensions of subjectivity, the body, and infrastructure to reach a unique understanding of intelligent systems informed by this metaphor of the black box as a site of temporal anomaly and ontological potential.



Zhang Wenxin (张文心), *The Inorganic Mysteries* (无机者的秘仪), 8K video animation with sound, 2019.

In *The Inorganic Mysteries* we wake up in a synthetic environment, peering out into a grey-toned digital cityscape through the eyes of the animation's protagonist as a moving scroll of mysterious rituals belonging to strange inorganic beings unfolds before us. There is a scene in which one of these beings touches the 'screen' between the protagonist/viewer and itself, and for a fleeting moment the inside/outside relation is inverted. Zhang then guides us through a progressively fractalised journey evoking what she

imagines a computer-generated intelligence may endure. *The Inorganic Mysteries* weaves together references from the ancient Greek Eleusinian Mysteries and the story of Chaos, the art of Buddhist mandalas, and meditation techniques from the Daoist classics, and instead of unpacking or demystifying the black box, the protagonist/viewer is simply made to traverse it. For Zhang, phenomenological experience carries more weight than functional capacity when it comes to the evaluation of an artificially intelligent being. If neural networks are approached merely as webs of matrix calculations, opaque vessels for inputs and outputs that forbid interrogation because of their immense complexity, the cave metaphor facilitates an approach to the black box via internal experience, and in *The Inorganic Mysteries* viewers are invited to share in this experience.

Recently, Zhang has been revisiting the Daoist idea of *tian shu* (天书). Often translated as ‘a book from the sky’, which is indeed its literal meaning, the term carries complex connotations in Chinese, where it refers to a sacred language as well as to illegible or nonsensical writing. A more accurate translation would perhaps be ‘calligraphy from the sky’ because it foregrounds the visual property of the medium, alluding to a kind of writing whose semantic power is located in its visual form. In a project that invites palaeographers and AI researchers to study patterns in data related to stalactites and ancient hieroglyphs, Zhang and her collaborators worked on designing an algorithm that would be capable of retrieving a holophrastic view from its training data.[5] Whereas large-scale natural language processing models such as GPT-4 imitate modern language systems by using probability predictions, this protolinguistic algorithm would be designed to mimic ancestral languages that were both texts and images, following the model of *tian shu*. Cross-modal learning and cross-medium representation would imbue it with the potential for chaotic yet interpretive output, a realm where meaning can unfold, in a way that is entirely distinct from that of the contemporary AI landscape, which is rife with image models and language models (and ways of plugging one into the other), but from which a model that can deal with image *as* text or text *as* image, falling outside the logic of algorithms, is noticeably absent.

Contemporary Media Studies views the invisibility or black boxing of infrastructures as something to be interrogated and, if possible, disrupted. For example, Jennifer Holt and Patrick Vonderau in their essay ‘“Where the Internet Lives”: Data Centers as Cloud Infrastructure’ point out,

summarising Bruno Latour, that ‘the mediating role of techniques is notoriously difficult to measure, at least as long as the machines run smoothly’, since ‘the more technology succeeds, the more opaque it becomes’.[6] In *The Marvelous Clouds: Toward a Philosophy of Elemental Media*, John Durham Peters calls upon us to broaden our understanding of what constitutes infrastructure across the nature/culture divide and to consider features of the natural world as technological interfaces and, echoing Marshall Mcluhan, underlines the ethical importance of making these ‘elemental media’ visible.[7] For all of these authors, failure of the interface can operate as an entry point, facilitating an ethical and epistemologically driven methodology of penetration/intrusion/inversion. The936 *shuiyue jinghua* metaphor, however, has more ambiguous implications, simultaneously highlighting the generative agency of infrastructures (water and mirror) and the ungraspable essence of the images that appear on its surface, so that the technology and its processes reveal themselves in a non-intrusive way, without being subjected to a drive to ‘get to the bottom of things’ (for there is nothing behind them to find).[8] Instead, the reflected images dissipate into a void upon the realisation of their intangible and elusive nature, a kind of knowledge that can only be attained by traversing the ethereal realm of the infra-thin ‘interface’ layer. An important element of *shuiyue jinghua*, concealed in its rhetoric, is that the lesson it seeks to convey invokes a bodily endeavour. Ancient mythic stories in China often depict the Sisyphean efforts of a hapless scholar in approaching the reflection of the moon or the flower and reaching out to touch what they take to be a real object only to be startled by the sudden revelation of its immateriality. Philosophically speaking, the ‘real’ moon and flower correspond to the morphing, hallucinatory, negative forms of reality, hence their reflections only serve as figures of potential emergence: realisations of the ‘unreal’.[9]

The traversal narrative staged in *The Inorganic Mysteries* also features in *Deep Simulator*, a 2020 game-based work by Shanghai and Berlin-based artist, activist, and programmer aaajiao. In a simulated world, across six levels distinguished by their use of computing power, the player passes through the Buddhist bardos of Life, Dreaming, Meditation, Dying, Dharmata, and Rebirth. The first three occur within one’s lifetime, while the second three describe states between life and death. Deep Wanderer, the protagonist of *Deep Simulator*, embarks upon a journey that requires them

to ‘mind the deep’. Unaware of the true nature of their existence, they are told by one of the key characters in the game that their presence is nothing but a ‘stream of wandering consciousness’, that ‘the body that [they] think [they] can see and feel is just an illusion in [their] mind’.

Totem
Body. The body that you think you can see and feel is just an illusion in your mind. Your current present being is nothing but a stream of wandering consciousness.
But someone told me that he saw that my body was illuminated by the fire. Space also told me that he saw something predicted on my body. What does it mean?
look at
one

Above: aaajiao, *Deep Simulator* (深渊模拟器), video game, 2020. Below: aaajiao, *A.I. Goooooooooogle infiltration* (人工智能 Goooooooooogle 渗透), video installation, 2019.

In an earlier video work called *A.I. Goooooooooogle infiltration (2019)*, which was installed on the ground floor of the former IBM building on Ernst-Reuter-Platz, aaajiao modifies images from AVA, an acronym for ‘Atomic Visual Actions’, a series of annotated moving image datasets developed by the Perception Team at Google AI with the end of goal of enabling AI to correctly identify human actions and gestures: ‘While maintaining AVA’s visual language, with its purple bounding boxes, he eliminates the desired information’—deleting the parts of the image that

contain actions or gestures.[10] This raises questions such as: What happens when what the algorithm detects is nothingness and—anticipating the themes of *Deep Simulator*—what does it mean to wander in the deep? The disappearance of what once filled the bounding boxes marks a return to a space of probability—what the black box stands for in Zhang Wenxin's work, a retreat from finite algorithmic truth. In *Deep Simulator*, furthermore, the disappearance of a stable body makes space for a nonlinear, informal, and unobjectified mode of being in a world run by algorithms. The cave is also present in *Deep Simulator*, connecting the different states of life and death. The bardos are the ultimate metaphor for latent space and the liminal state between current instantiations of AI and its future dream: consciousness. In the Daoist view, the human body contains a hologram of the structure of the universe, hence, reaching outwards is equivalent to looking inwards. In *Deep Simulator*, the Deep Wanderer can switch Dharma vessels between the 'Floating Wileys' (inwards) that generate dreams and the 'White Glove' (outwards) that dissipates all illusion—illusion being precisely what lies behind the mirror and beneath the surface of the water. Upon finally completing the game the player controlling Deep Wanderer emerges in a simulation of themselves playing *Deep Simulator* in a room evocative of the very gallery space that they have been standing in all along.

He Zike (贺子珂), *E-dream: We'll stay, forever, in this way* (电子梦:我们将如此继续), real-time web program, unlimited duration, 2018.

Even for artworks composed using realtime data, the generated scenes seem to unfold in a subtle temporality all of their own. He Zike's work *E-dream: We'll stay, forever, in this way* captures realtime news data, fragments it, and recomposes it using a fine-tuned GPT-2 model into the ongoing story of

‘Jack’ and ‘Rose’, two protagonists eternally trapped within the confines of the screen.[11] Over time, He’s creation has metamorphosed into a poignant reflection on reality itself. In the midst of the pandemic, in response to the contents of mainstream international news, the daily lives of Jack and Rose were inundated with keywords such as ‘medical’, ‘masks’, and ‘virus’, seamlessly integrated into the narrative by the relentless operation of the algorithm. Since it is based on realtime data processing, the story itself has no beginning or end, and any encounter with it can only be temporary. Its deliberate resistance to a comprehensive narrative structure imbues it with a kind of timelessness, eschewing climactic moments, meticulously crafted settings, and the notions of surplus or scarcity. However, it is in the occurrence of its frequent linguistic glitches that one may discern the subtle ripples in the fabric of its otherwise reportage-like realism. Drawing inspiration from the annals of automatic writing, Charles Bernstein’s echopoetics, and the exquisite corpse technique pioneered by the French Surrealists, the artist meticulously explores the phenomenon of apophenia (the inclination to establish connections between unrelated elements) within the realm of pattern recognition. Apophenia further adds to the poetic dimension of *E-dream*, rendering it simultaneously plausible and enigmatic, primarily relying on the interpretative realm that intervenes between the viewer and the screen. Much like how *shuiyue jinghua* initially addressed the hermeneutic space encompassing poetry and the inherent agency of the reflected image, the work evokes similar affective responses from its viewers. Amongst the thousands of lines generated, one of them reads: ‘Rose passed away, people found coronavirus in her palms.’

As the involvement of Chinese artists in AI-related work continues to expand, a broader spectrum of artistic practices has emerged, providing oblique entry points into the developing AI landscape. The works discussed in this text seek to show that the potential of artificial intelligence relies neither on a clumsy imitation of our own minds, nor on pure technical optimisation; it is closer to a holographic inheritance of human life experiences, which include the illusory, the confusing, and the unpredictable, alongside momentary tangents with seemingly distant, ancient echoes. *Shuiyue jinghua* is not just a philosophical parable but an epistemological instrument that can reveal the unstable, paradoxical, poetic, and transtemporal dimensions of simulated intelligences and their virtual worlds. The mediating interface, the ‘water’ or ‘mirror’—the artistic

medium—is a portal through which plural narratives unencumbered by the constraints of a singular technological perspective or promise can be traversed. It also creates a leak in the hegemonic regime of operational, automatic, and relentless computing and its ideal of ‘smooth’ reality. As Matthew Fuller and Andrew Goffey put it in *Evil Media*, a leak facilitates ‘an attempt to capture and mobilise the dynamics of unintended consequences’—‘the double agent, confusion, [rendered] fruitful’.[12] And finally, in addition, *shuiyue jinghua* leads to ‘sudden insight’, *dunwu* (顿悟), in a Buddhist sense. In *Master Jiumoluo’s Mahayana Great Doctrine* (鸠摩罗什法师大乘大义), written during the Dong Jing dynasty by the monk Huiyuan (慧远), conformity to the Dharma body is described as ‘moon in the water’, visible yet intangible, a superposition of *se* (色) and *wuse* (无色)—the aggregation of form and its absence.

Shi Zheng’s ghostly machine visions, Zhang Wenxin’s black-box traveller, aaajiao’s interzone of artificial agents, and He Zike’s neverending electronic dream, along with the work of many other AI artists in China, collectively outline an artificial world which holds countless reflections of fragments of our world. If the hype surrounding artificial intelligence is about demonstrating its infallibility, accelerating its ability to mimic human consciousness, and generating the ‘real’ moon and flower, the artistic approaches examined in this text are, on the contrary, more interested in exploring the water and the mirror—the media, interfaces, and portals by means of which we can interrogate the nature of ‘reality’ itself.

Notes

1. Foundation models differ from their predecessors owing to the extremely large size of the datasets they are trained on, and the fact that their training is self-supervised and utilises unlabelled data. The breadth of the data they are trained on makes it easy to adapt the same model to a wide variety of different tasks. Owing to their large scale, they display various emergent properties that elude our current understanding, hence their black-boxed nature. The term was coined by the Stanford Institute for Human-Centered Artificial Intelligence (HAI) Center for Research on Foundation Models (CRFM) in 1. See R. Bommasani et al. ‘On the Opportunities and Risks of

Foundation Models', Stanford University HAI CRFM, <https://crfm.stanford.edu/report.html>.
2. See An, Bo, 'Between the Human and the Machinic: Qian Xuesen and AI', in this volume.
3. 'Smooth' is a technical term for the state in which the temporal lag or latency in a simulation is below the threshold of human perception. Smoothness is achieved when a human user feels that there is no difference between a simulation and 'real life'.
4. Q. Meillassoux, *After Finitude: An Essay on the Necessity of Contingency*, tr. R. Brassier (London: Bloomsbury, 2009), 64.
5. Zhang Wenxin (张文心), *Proto Language* (work under development).
6. J. Holt and P. Vonderau '"Where the Internet Lives": Data Centers as Cloud Infrastructure', in *Signal Traffic: Critical Studies of Media Infrastructures*, ed. L. Parks and N. Starosielski (Urbana, Chicago, and Springfield, IL: University of Illinois Press, 2015), 74.
7. See J.D. Peters, *The Marvelous Clouds: Toward a Philosophy of Elemental Media* (Chicago: University of Chicago Press, 2015).
8. Ibid., 34.
9. The story can also be read as a means of inviting us to realise that we ourselves and the 'reality' we feel we are immersed in are reflections too, just like the moon and the flower.
10. Efremidis Gallery Press Release, September 2019, <https://eventstructure.com/A-I-Gooooooooogle-infiltration>.
11. He Zike (贺子珂), *E-dream: We'll stay, forever, in this way* (电子梦:我们将如此继续), real-time web program, unlimited duration, 2018.
12. M. Fuller and A. Goffey, *Evil Media* (Cambridge, MA and London: MIT Press, 2012), 103.

Suzanne Livingston

Curatorial Escapology Reflections on AI: More than Human

There are more than enough terrors to be found there; but such terrors are not all there is to the outside.

MARK FISHER[1]

There are few topics as intangible and indefinable as artificial intelligence. Its conceptual roots can be traced back to the dawn of time; its future takes us far beyond the known world. It is shot through with polemic and wild speculation: what it is and what it is not is the subject of continual debate. Giving physical form to this abstract subject was a curatorial challenge we faced while working on a major landmark exhibition surveying the history of AI for the Barbican Centre in London.

The exhibition, entitled *AI: More than Human*, opened in May 2019, after a three-year-long planning and production process.[2] It offered the public—ninety thousand visitors in its first three months—a comprehensive overview of the history, social effects, and potential future of AI. It continues to tour across continents and cultures. At the time of its conception, it coincided with a season of AI shows across Europe which focused on art, creativity and AI, or the ethical questions that the technology raises. *AI: More than Human* was different for its scale, its interdisciplinarity, its transculturalism, and for its philosophical enthusiasm about the future. Its broad lens was essential for challenging the Western AI narrative and expanding common perspectives on AI, pushing some viewers out of their comfort zones. It was experimental in the connections it made, fearless about the science, and (something that is unusual for the field) was curated by a core team of women.

Despite the indeterminacy of the subject matter, the show retained a clear point of view. The crux was the philosophical and technological borderland between Europe and the US, Japan, and, as much as was practically possible, China. In terms of the development of AI, there is a vast history of interconnections between these cultures, but there are also dominant, often

superficial narratives that locate specific attitudes and relationships to the technology in these different parts of the world—one set of relations between capitalism, technology, and consumerism in the West; another between government, authoritarianism, and population in China; another between nationhood, youth culture, and innovation in Japan. The intention of this show was to ensure that the Western AI narrative was seen as just one narrative among others, most particularly for those viewing the exhibition in a Western context. As co-curator, my collaboration with Maholo Uchida, then Chief Curator at Miraikan in Tokyo, was essential to this endeavour. To her international experience in museums and technology, I added my background in philosophy and my experience in strategy consulting for the technology world, drawn from time in China, Japan, Europe, and the US. Together with the Barbican team we traded our knowledge and understanding of AI narratives in these different parts of the world.

As a result of this transcultural conversation, and to the surprise of many visitors, the exhibition set off from a point five hundred years in the past, in deeply philosophical territory. The first section explored Japanese Shinto, mystical Judaism, sixteenth-century alchemy, and early Gothic literature, providing very different perspectives on the possibility and desirability of the inanimate coming to life. The second section considered early efforts to recreate the brain as a series of computational processes, including the work of Charles Babbage, Ada Lovelace, and Alan Turing, alongside the roots of such efforts in early calculation tools from Europe, China, and Japan. Crucial to this part of the journey through the exhibition was the historical shift from symbolic AI, which uses human readable representations of problems, to sub-symbolic AI (or connectionist AI) which draws directly upon emergent biological processes. Sub-symbolic AI—which took off in the 1980s—does not use representational logic. It teaches itself, and is far better at dealing with uncertainty than symbolic AI, but it is also harder to understand how it reaches a solution to a problem. This predicament was demonstrated in the centrepiece of the second section—a rich, interactive exhibit documenting the history of DeepMind’s AI Go player AlphaGo, culminating in the story of its triumph against 9 dan Master Lee Sedol.

The third section looked at the presence of AI in daily life, in obvious and less obvious ways. It profiled Regina Barzilay’s MIT work on using AI to detect the potential for breast cancer in patients years in advance of other

commonly used diagnostic technologies; Amnesty International's arms trade tracking systems, which predict the flow of illegal arms around the world; food technologies producing crops in the harshest and most inhospitable environments; AI toys from Shenzhen; deepfakes rewriting the speeches of Barack Obama; Joy Buolamwini's work on the problem of racial and gender bias in data sets; Nao Tokui's work in AI music composition, and Google Jigsaw's technology designed to weed out bullying in online chat forums, developed for *The New York Times*. Here, bigger questions of privacy, control, coercion, and collusion were confronted. For every AI innovation pointed toward a more equitable world, there is a counter-innovation demonstrating how well the technology can lend itself to the manipulation of power. The final section delved into a future where artificial intelligence becomes molecular, in the deep terrain of artificial life. Here, living systems, and sometimes living beings, are conceived from scratch, but also reconceived in terms of their function and capability, while brains are decoupled from bodies and no longer need to die with them. This part of the exhibition covered gene editing, mind uploading, cryonics, and new forms of synthetic nature. It was the conclusion of a journey exploring the human desire to build an artificial mind, animate an artificial body, and even rework the rules of nature itself.

The exhibition did not seek to delimit AI or impose a human-centred story upon it. The goal was not to confirm humanity's place in the world but rather to expose human limits and then to explore the questions and terrains such an approach opens up. This is a hard journey to take an audience on—it can be challenging and discomforting. Reactions to the London show were an interesting bellwether of attitudes to the technology at that moment in time. Many visitors enjoyed the opening of new conceptual doors that the show inspired but there were less open-minded responses too, and they can be instructive to consider. In general, they were characterised by a number of different convictions: that AI will stay under human control; that it is only a tool; that it is merely overblown hype; that its single-minded goal is human destruction; that it is certainly not creative. These are all reactions that need to be engaged with, especially in public space. They reveal an anxiety that cannot be ignored. From my perspective, they may also be representative of a particular kind of closed-loop anthropocentric thinking —precisely the kind that will be outpaced by AI, leaving its proponents, almost by way of self-fulfilling prophecy, exposed to the widening gap

between artificial intelligence and the human ability to engage and keep up. The only thing that is certain about AI is how unknowable it is. It is working at scales, at speeds, and in dimensions that differ dramatically from those of humans, from whose perspective very few of its internal operations yield themselves up to decipherability. Neuroscientist Daniel Kahneman is sanguine about the future it points towards: ‘Clearly AI is going to win. It’s not even close. How people are going to adjust is a fascinating problem.’[3] As he explores the conceptual issues which hold people back from engaging with AI as it actually is, he points out the deep-seated human preference for natural systems over abstract, artificial ones. He uses COVID-19 vaccination as an example: ‘People are willing to take far, far fewer risks when they face vaccination than when they face the disease.’[4] This perceived gap between the natural and the artificial—or the human and the machinic—only doubles down on the kind of thinking that will be outpaced by AI. To escape this dismal prognosis, there is an urgent need to step into what Mark Fisher, in the phrase quoted in the epigraph to this essay, calls ‘the outside’.

Seeping Realms

From a curatorial perspective, the topic of968 AI is full of invitations to explore the outside. My intention here is to use the term ‘outside’ in its philosophical sense, denoting the vast and sometimes frightening realms of reality beyond human perception, cognition, or experience. Such realms are integral to the creation of AI, composed as it is of layer upon layer—mathematical, technological, conceptual—of philosophical unknowns. Deep learning embodies a historical shift from linear mathematics and fixed outcomes to finding patterns which we ourselves cannot see in vast quantities of unlabelled data. For it to work, it must reach beyond human guidance and human reasoning. Whilst we may hope to wrestle its unknown realms down to earth and bring AI under epistemological control, AI’s relation to the unknown has also been the driver of its profound attraction for researchers and experimentalists for centuries. Integral unknowability is its engine and adrenaline. As much as it is a closed book, AI provides access to something greater. It takes us beyond our abilities and our constraints, and it brings to the surface connections in data and in the world which otherwise would remain imperceptible to us. It therefore offers a

conduit to the outside which can be accessed via a method of ‘escapology’[5] —a means by which we can alienate ourselves from ourselves and open up radically new exploratory spaces in which we have no easy footing.

AI: More than Human aimed to open up a number of escapological hatches and chutes. It was an exercise in revealing hidden pathways between worlds, technologies, and ideas, with a curatorial methodology designed to present a deep perspective on AI in which everything is more entwined than familiar narratives suggest. What at first may look rational can often turn out to be irrational, what might seem human can often turn out to be something else, and things we have assumed to be quintessentially Western often have their roots in other cultures and traditions. One example featured in the show was a Victorian replica of Leibniz’s 1671 calculating machine. A golden box reminiscent of the highest form of watchmaking, distinctive for its geared hand dial, and unprecedented in its ability to perform the four key arithmetical functions (addition, subtraction, division, and multiplication), Leibniz’s machine reflected the philosopher’s near-obsessive ambition to systematise and classify the world around him. In a speculation that appears to be a perfect precursor to the intensely quantified culture of today, he wrote,

> we should be able to reason in metaphysics and morals in much the same way as in geometry and analysis. If controversies were to arise, there would be no more need of disputation between two philosophers than between two accountants. For it would suffice to take their pencils in their hands, to sit down to their slates, and to say to each other...: Let us calculate.[6]

Distilling philosophical concepts into mathematical equations was more than just a reductive, unidirectional idea. For Leibniz, it ran the other way too, opening a doorway to subtle, numinous worlds. ‘The Step Reckoner’, as the calculating machine was known, was an important development on the journey toward one of Leibniz’s biggest breakthroughs—the modern binary system, built on the digits 0 and 1. Because it is able to rapidly navigate vast systems of numbers thanks to its core binary function correlating to the Boolean logic (‘on’ and ‘off’ states) that all computers understand, the binary system has since become the universal language of computing, and this will be the case until true quantum computation is reached.

Leibniz did not however create the modern binary system with computing as his primary motivation, rather, his work was driven by religious belief. He saw his system as lending credence to his Christian ideals, as an

explanation of how God generated the universe out of nothing. Just as he pursued his interest in this most rational ordering of the world, he simultaneously delved deep into the most inconceivable domain of all. Much of his work was carried out within the narrow philosophical and religious frameworks of the Western Enlightenment, but it stretched far beyond them too. The faith he found in his binary system was bolstered not just by Christian principles but by ancient Chinese philosophy. As is revealed in the title of his most important paper on the subject, 'Explanation of Binary Arithmetic, Which Uses Only the Digits 0 and 1, with Some Remarks on Its Usefulness, and on the Light It Throws on the Ancient Chinese Figures of Fuxi',[7] the foundation for modern binary was the *Yijing*, one of the oldest classical Chinese texts, a pillar of Confucian, Daoist, and Chinese Buddhist cosmologies. Through his passion for China and its culture, communicated to him by travelling Jesuit missionaries, Leibniz found an affinity with an ancient belief system that has, for centuries, been used as a divination tool to access the universe's deeper patterns—not to straightforwardly predict the future but rather to touch upon the multitude of adjacent, virtual potentialities and realities that lies beyond the immediate experience of a single point in time. This itself became a notion that would be repeated again and again throughout Leibniz's work on potentiality and actuality, most notably in the *Monadology*.[8] At its core, the *Yijing* is a highly technical system built from simple binary elements (the tossing of a two-sided coin, the broken and unbroken lines of the hexagrams) which provides for its users access to the workings of the cosmos. For Leibniz this insight into a logic far off the human map was the affirmation he needed to believe in the potential of his binary system. This most 'rational' of numerical systems was hardly conventionally rational at all, and yet it was also deeply technical. Like much of AI's technological history, the modern binary system is the product of a vast silting of mythos and techne, sacred belief and concrete practice, whose real intercultural complexity is very easily forgotten.

The profound entanglement of rational machinic processes with cosmological systems is more than historical. In 2016, DeepMind's extraordinary AI Go player AlphaGo competed against Lee Sedol in Seoul, South Korea, with over million people watching worldwide. Over the course of five dramatic games, AlphaGo won four of the five through a series of often bizarre moves. It never mattered to AlphaGo how much it

won by, only that it won. And win it did, wearing down and confusing its opponent over and over again until his final resignation. In *AI: More than Human*, the focal point of this extraordinary story was ‘move 37’. Move 37 was a perfect demonstration of the wild potential of nonhuman intelligence, and has since triggered an enormous amount of commentary and analysis. As Fan Hui, himself an outstanding Go player and a key collaborator in the technological development of AlphaGo, commented: ‘[Move 37] is not a human move. I’ve never seen a human play this move.’[9] Move 37 also set the scene for the almost as famous move 78, Lee Sedol’s retort, now commonly referred to as ‘God’s Move’—a product of so-called divine inspiration. Move 78 also broke new ground in its radical challenge to playing convention. Together, these vital manoeuvres took the game into new territory, establishing a playing style that human Go players have been honing ever since.

What cannot be underestimated is the importance of Go itself in this moment of extreme technological breakthrough. A deceptively simple game of competing black and white pebbles in which each opponent must enclose and capture the pebbles of the other, Go is a cultural phenomenon rooted as much in cosmology as in the *Yijing*, and much has been written about its sacred roots in shamanism, astrology, and divination. As game historian Peter Shotwell puts it, the universe can be made to speak through Go.[10] The game is said to offer more moves than there are atoms in the universe, and in its combination of precision, limitlessness, and immunity to epistemological circumscription, it has proven itself to be the most fitting tool for sharpening the power of artificial intelligence. Synthetic minds are meticulously honed in a virtual space of play whose possibilities may never be exhausted. According to Go expert Frank Lantz, ‘people who play Go don’t just want to understand Go, they want to understand what understanding is’.[11] And understanding understanding is always a journey to the outside.

Alien Rhythms

> *A fish trap is there for the fish. When you get the fish, you forget the trap. A snare is there for the rabbits. When you get the rabbit, you forget the snare. Words are there for the intent. When you get the intent, you forget the words. Where can I find a man who has forgotten words, so I can have a word with him?*
>
> ZHUANGZI[12]

‘Explainable AI’ is a relatively new discipline, and as the name suggests its intention is to counter the opacity of the deep learning techniques that predominate today. There are calls for such explainability, or interpretability, to become a crucial and legal requirement in vital domains of AI such as medicine, finance, defence, and law. Doctors want to understand which aspects of a patient’s physical state led to the conclusions reached by a diagnostic AI, and bank customers want to understand which aspects of their loan application led to its being refused by a financing AI. Attempts have been made to fulfil these demands, but they often meet with failure. According to a recent study of the behaviour and readability of AI, ‘decision trees [that] have an excellent degree of explainability […] exhibit [the] worst prediction accuracy’ while ‘[i]n the other extreme, deep learning methods are better in predictive capacity than any other learning methods but they are [the] least likely to be explicable’.[13] In other words, the desire to develop the most effective AI runs counter to the effort to understand it. The authors argue that progress will be made in enabling deep learning to share aspects of its methodology, in some cases simply in verbal terms, but it will only ever be a partial and distorted view. Two techniques are in development to achieve this. The first is ‘transparency design’ whereby the program reveals a snapshot of its functionality to the user. The second is ‘post hoc explanation’ in which the AI model effectively rationalises and explains itself. They are both seen as vital steps toward ‘human in the loop’ AI.

Increasing AI transparency is integral to the work of computer scientist Regina Barzilay, which featured in the third section of *AI: More than Human*. Barzilay is responsible for developing a deep learning imaging program capable of recognising clinical signs of breast cancer years ahead of other existing techniques. It is important to Barzilay that the systems she is building reveal their decision making—indeed, this is essential if they are to be truly medically useful. To bring more of what is going on inside her program to the surface, she has found a way to have them extrude snippets of text that indicate the patterns they have discovered—a snapshot of AI thinking in action.[14] Although it does not provide the full picture, this is one way of trying to decrypt the flow of decision making within deep learning processes. Barzilay’s research is revealing vital flashes of insight, but also suggests that the black box is bigger than our ability to decipher it.

In the research field of artificial life, which occupied the final section of the exhibition, these unfathomable depths are not feared but relished. A-Life is dedicated to understanding complex processes in the biological world through highly sophisticated simulations. For experts such as Professor Takashi Ikegami, whose ethereal robot Alter 3 entranced audiences visiting *AI: More than Human*, the development of artificial intelligence is merely a side project in the field's larger mission to understand the building blocks of life itself.[15] Unlike the goal-setting nature of artificial intelligence, which points itself at practical problems in the world, artificial life does not seek to stand apart from the processes it studies. It siphons and channels them rather than capturing or even attempting to explain them. From the perspective of molecular biologist and neuroscientist Jeff Lichtman, human minds bring with them very human problems that must be carefully navigated in the field of A-Life: 'Language itself is a fundamentally linear process, where one idea leads to the next. But if the thing you're trying to describe has a million things happening simultaneously, language is not the right tool.'[16]

One of the final exhibits in *AI: More than Human* was Jeff W. Lichtman's Brainbow project, a study of the brain developed over the past fifteen years, which shows that, although the human brain has evaded understanding time and time again, the day we will be able to produce a complete map of its neurons is not as far off as we once thought. Whilst the field of neuroscience has been able to identify the function and location of regions of the brain, the interaction and interdependency of these regions is largely unknown. In the 1970s it took more than ten years to map the neural circuitry of the *caenorhabditis elegans*, a one-millimetre-long worm.[17] This produced the first complete wiring map or 'connectome' of an animal brain, detailed across more than three hundred pages of paper documentation. *Caenorhabditis elegans* has fewer than four hundred neurons to plot, while human brains have eighty-six billion. Luckily, Lichtman has already started on an intermediary project: a mouse brain. Lichtman's mapping process begins by slicing the mouse brain into thirty-nanometre-thick sections. These are then fed into an electron microscope and blasted with electrons to produce pixelated images that start to reveal the patterns and interactions of axons, dendrites, and synapses. Further work is required to move from these hazy two-dimensional images to the rich volumetric wiring maps for which Lichtman is known. Previously this was done manually by willing students,

but more recently it has been done by AI image recognition and processing software. Processing brain datasets was a task that previously took years to complete. Now it can be achieved in hours. AI has sped up the process so much that fully mapped brains are finally on the horizon. In a perfect feedback loop, AI is automating processes which help us understand animal brains, furthering our understanding of our own brain so that we might develop artificial brains in turn.

Viewed through a speculative lens, Lichtman's work has some extraordinary implications—for instance, brains living on servers, independent of the bodies they were born in, for use after death or in tandem with other living systems. Lichtman himself feels wary about such possibilities and he reminds us that these wiring maps, as phenomenal as they are, are only representations. They may help us solve some of the most difficult scientific problems, but ultimately they cannot reveal everything about the dynamic reality of thought:

> That's what the human brain is. It's millions of things happening simultaneously among different types of cells, neuromodulators, genetic components, things from the outside. There's no point when you can suddenly say, 'I now understand the brain…'.[18]

The map is not the territory—especially when the territory itself is a complex ever-changing field of interrelating rhythms and milieus.

Spiral Matter

> *If AI models cannot be reduced to human terms of reference, perhaps human thought can be expanded to comprehend computational terms of reference.*
>
> VENKATESH RAO[19]

In the domains of artificial life and artificial intelligence, the rational and irrational, and the representational and non-representational, are entwined. So much of what is outside human perception and beyond human cognitive grasp is constantly creeping in. Indeterminacy is everywhere. In contemporary Western culture, there is a drive to resist the forces of AI, but at the same time, we can't seem to say no to its allure. It is the lodestone of culture—the doorway to something far greater that has fascinated and unsettled humans for centuries. The Gothic literary tradition, which featured at the beginning of *AI: More than Human*, is dedicated to exploring humanity's place in the world and its destruction by forces that cannot be understood or foreseen. Time and time again, the Gothic tradition returns to

themes that allow it to experiment with ways of defamiliarising the human. In many stories, an interest in awakening the dead, animating the inanimate, or tinkering with the categories that secure humanity's place in the world becomes a fast route to a character's demise. E.T.A. Hoffman's short story 'The Sandman',[20] which is exemplary of the genre, occupied a central place in the exhibition. 'The Sandman' is a tale of possession, hallucination, breakdown, and mistaken identity born of a love affair between the protagonist, Nathaniel, and an automaton. Names and identities blur and merge, characters (including Nathaniel) become strangers to themselves and one another. For Sigmund Freud, whose subsequent work bestowed great cultural significance on the tale, Hoffmann's story was an embodiment of his theory of the uncanny—that eerie feeling we experience when something inanimate becomes too lifelike. Freud explained the sensation as a 'regression to a time when the ego was not yet sharply differentiated from the external world and from other persons', and as an involuntary return to a 'primitive', 'animistic stage' where the boundaries between the living and the dead, ego and id, have not yet emerged.[21] The uncanny is what triggers the disassembling of identities and the psychic unravelling of Nathaniel—the automaton is his undoing. The parallels between the thematic obsessions of the Gothic literary tradition and artificial intelligence are obvious: both challenge the definition of the human by troubling the boundaries between living and dead, animate and inanimate.

For Venkatesh Rao, the demise of the comfortable category of the 'human' is a relief, finally marking the end of an era of rigid, ring-fenced, anthropocentrism. Thanks to what Rao calls 'superhistory' rather than 'superintelligence'—he maintains that 'modern AI is better understood as AT—"Artificial Time"'—we might finally come to see past 'individual-human-mind bottlenecks', the 'identity insecurities of various intellectual traditions', and 'our own failing senses and increasingly limited experiences' by means of AI's 'superhistorical' prostheses.[22] Rao gives the example of a GPT-3 writing assistant imbued with centuries of knowledge acquired by means of gigantic data sets which dwarf the storage capacity of any individual human mind. This technology, based as it is on 'vastly more text than [one] could read in a thousand years', gives the user the ability to write 'like a 6-year-old rather than a 46-year-old'. As well as bringing domains that far exceed individual human cognitive capacity or experience into our purview, AI is also recontextualising humanity's place in history. In

Artificial Time, our place shrinks. The implications of this play out in interesting ways in the context of the museum. According to Rao, '[t]hese old institutions were built as extensions of ourselves, reproducing and even intensifying all the extreme repetitiveness and redundancy of our own lives', but they were unavoidably 'shaped by costly storage technology'. In order to build them, 'individual human brains had to digest manageable chunks of data and turn them into informational bricks' amenable to presentation and consumption. With algorithms working on the scale of history and drawing on data not limited to individual human experience, however, superhistorical AI 'blows past our own limits and discards our past in ways traditional institutions were simply incapable of', destroying 'egoistic modes of cultural production, preservation and transmission'.[23]

Not only does the human perspective on human 1history have its limits, but human history itself is tiny in relation to everything else. Beyond the human mind, there is an infinity of cognitive and experiential possibilities, and AI is already there, interfacing with worlds beyond our world and pulling the familiar apart. To think with AI, rather than attempting to contain it, is to challenge the borders between mind and body, nature and culture, rational and irrational, representable and unrepresentable, calculable and incalculable. None of these distinctions is useful in and of itself, and no one term should be the backdrop to any other—they are all enmeshed, silted up in messy structures continually in motion. To tune into the world as it really is requires constant movement—to go beyond categories, to keep shifting perspective across cultures, across species, across sexes.

Far more than just an exhibition about artificial intelligence, *AI: More than Human* unfolded in dialogue with the outside logic of AI to create an escapological opportunity—exteriorising and defamiliarising its subject matter, undoing rather than retelling the timeworn story of human centrality. Standing outside the Western, humanist viewpoint, and allowing AI's strange logic to do its work, the exhibition ciphered stories from the edges and the in-betweens, from inside machines and amongst materials. This is a means of navigation rarely explored in museums, one in which the human viewer is not the ordering gaze but instead a transmitter of an object's own power as it reveals histories and futures that extend beyond human use. Escapology as a curatorial methodology is the search for new hooks and webs that radically shift vantage points and orientation. It means starting

without an identifiable destination and using unknowability to escape our constraints, bringing whole new dimensions to the fore—but in a way that requires us to rethink what we take for granted about ourselves.

Daoism—an ambient presence throughout the exhibition, threaded through the history of binary code and the *Yijing*, Go and AlphaGo—offers a philosophical model for the escapological rules of play. It is a system which has no constructed laws beyond those of the cosmos itself, no central deity, no fixed destination. It is pure, exploratory, generative becoming. In its modus operandi, it shares a great deal with advanced forms of AI such as DeepMind's game playing 'MuZero', which, unsupervised, can learn rules within minutes and acquire posthuman playing skills within hours. MuZero could be said to demonstrate Daoist-like intelligence in the way it senses environmental cues and finds feedback patterns and affinities, developing over time powerful synchronicities and entering states of flow. At the bleeding edge of deep learning, MuZero navigates its environment with Daoist sense organs.

According to the words of the *Zuangzhi*, the best life and the deepest intelligence comes from seeing and feeling things in the present in their wider contexts, in tune with their knock-on effects, from all perspectives simultaneously, and as a connected system. Only AI can help us unlock insight on such a massive scale. This is a way of being that does not emerge overnight, but humans have been holding onto the old guardrails for far too long. New ways of approaching the cosmos will reveal themselves, and the Western narratives upon which we too often rely will become no more than small-town thinking. It is time to let go. As my great, late friend and theorist of the outside, Mark Fisher, once said (intending to give me career advice but really espousing a principle for living life): 'When you have become an insider, it is time to move on…'.

Notes

1. M. Fisher, *The Weird and the Eerie* (London: Zero, 2024), 9.
2. Barbican International Enterprises, *AI: More than Human*, Barbican Centre, London, May–August 2019.
3. T. Adams, 'Clearly AI is Going to Win. How People are Going to Adjust is a Fascinating Problem', interview with Daniel Kahneman, *The Guardian*,

16 May 2021, <https://www.theguardian.com/books/2021/may/16/daniel-kahneman-clearly-ai-is-going-to-win-how-people-are-going-to-adjust-is-a-fascinating-problem-thinking-fast-and-slow>.
4. Ibid.
5. See B. Singleton, ‘Maximum Jailbreak’, in R. Mackay and A. Avanessian (eds.), *#Accelerate: The Accelerationist Reader* (Falmouth: Urbanomic, 2014), 489–507.
6. G.W. Leibniz, quoted in B. Russell, *A Critical Exposition of the Philosophy of Leibniz* (Cambridge: Cambridge University Press, 1900), 169–170.
7. G.W. Leibniz, in *Leibniz on Binary: The Invention of Computer Arithmetic*, ed. and tr. L. Strickland and H. Lewis, (Cambridge, MA: MIT Press, 2022), 189–98.
8. G.W. Leibniz, *Monadology*, tr. L Strickland (Edinburgh: Edinburgh University Press, 2014).
9. Quoted in the documentary *AlphaGo*, dir. G. Kohs, 2017.
10. P. Shotwell, ‘The Game of Go: Speculations on its Origins and Symbolism in Ancient China’, 2008, <https://vdocuments.site/the-game-of-go-speculations-on-its-origins-and-symbolism-likely-to-have-derived.html>, 45. (An earlier version was published as ‘Speculations on the Origin of Go’, in *The Go Player’s Almanac 1*, ed. R. Bozulich, [Chigasaki: Kiseido, 2001]).
11. Quoted in *AlphaGo*, dir. G. Kohs, 2017.
12. Zhuangzi, *The Complete Writings*, tr. B. Ziporyn (Indianopolis: Hackett, 2020), 224.
13. F. Xu, H. Uszkoreit, Y. Du, W. Fan. D. Zhao, and J. Zhu, ‘Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges’, *Natural Language Processing and Chinese Computing* LNCS 11839:51 (2019): 2–3.
14. See W. Knight, ‘The Dark Secret at the Heart of AI’, *MIT Technology Review*, 11 April 2017 <https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/>.
15. S. Livingston, ‘Between Man and Machine’, interview with Takashi Ikegami, Daiwa Foundation, 2 December 2021, <https://www.youtube.com/watch?v=3NGVIcK1KjE>.

16. Quoted in G. Guitchounts, ‘An Existential Crisis in Neuroscience’, *Nautilus*, 23 January 2020, <https://nautil.us/an-existential-crisis-in-neuroscience-237680/>.
17. E. Landhuis, ‘Probing Fine-scale Connections in the Brain’, *Nature*, 19 October 2020, <https://www.nature.com/articles/d41586-020-02947-5>.
18. Quoted in Guitchounts, ‘An Existential Crisis in Neuroscience’.
19. V. Rao, ‘Superhistory, Not Superintelligence’, *Ribbonfarm*, 11 May 2021, <https://studio.ribbonfarm.com/p/superhistory-not-superintelligence>.
20. E.T.A. Hoffman, ‘The Sandman’, tr. P. Wortsman (London: Penguin, 2016).
21. S. Freud, *The Uncanny in Art and Literature*, (London: Penguin, 1990), 358, 363.
22. Rao, ‘Superhistory, Not Superintelligence’.
23. Ibid.

Xin Wang

Machine Envy

Плохо ваше дело! По-видимому, у вас образовалась душа [You are in trouble! It looks like you have a soul].

YEVGENY ZAMYATIN, *WE* (1924)

The amount of data involved in managing society at all levels is increasing at terrific speed, as is the quantity of justifications for decisions. The decisions that will shape our society in the future will have to be arrived at, developed, and carried out through technology. But they must be based on the same criteria of respect and appreciation for human capacities, freedom, and responsibility that prevail in art. To paraphrase what Tristan Tzara said about Dada: 'No one can escape from the machine, only the machine can enable you to escape from destiny.'

K.G. PONTUS HÚLTEN, *THE MACHINE AS SEEN AT THE END OF THE MECHANICAL AGE* (1968)

请控制灵魂对自由的渴望，不要开窗歌唱 [Please curb the soul's thirst for freedom and refrain from singing from open windows].

DRONE MESSAGE TO A RESIDENTIAL DISTRICT DURING MASS QUARANTINE IN SHANGHAI (2022)

Bots

Since around 2018, the social media platform Weibo has seen a curious rise in bots, by which I don't mean automated zombie accounts for hire, or coordinated troll armies, but rather humans simulating bots by creating accounts with 'bot-sounding' handles.[1] Some are innocuous and fun, like 'cat bots'; others serve specific subcultures as quasi-archives, like the 'Millennium bot', or deliver quotable gems, like the 'French literature bot'. Then there are accounts like the 'Soviet Jokes bot' or the 'Lu Xun bot', which have now, after several rounds of censorship and resurrection, largely disappeared—save for scant remaining search results—likely owing to coded political expression in their output. Compared to their Twitter counterparts, which often similarly employ algorithms to autogenerate content, the nuance here feels exquisitely culturally mediated. The combination of Chinese characters with the English word 'bot' in the handles creates a unique dissonance. It escapes no one that these accounts

are managed by humans, since the content often relates to current events, which begs the question: What makes this performative simulation of machines and algorithms so appealing? And what makes it culturally specific?

‘Lu Xun bot’ was named after one of modern China’s most formidable and incandescent writers (1881–1936), whose trenchant critique of the ills of imperialism and Confucian conservatism continues to resound in national textbooks and public discourse alike. His cultural and political thoughts have been widely recited and, more crucially, promoted by the authorities as representative of their central values. One may argue that he constitutes an essential dataset on which generations of modern Chinese people have been trained, which makes Lu Xun’s words uniquely potent and thorny when deployed to critique the social ills of today. Examples range from his feminist quips on slut-shaming and calling out feudal values as anti-human (or, cannibalistic, as he likes to put it), to his descriptions of the general apathy he observed in Chinese society at moments of crisis: ‘the joys and sorrows of humans are not connected after all’—a sentence that has been repeatedly evoked to address increasing gaps and lacerations across social strata.[2] Coded yet performative, ‘Lu Xun bot’ exemplifies a uniquely creative form of cultural and political expression within Chinese cyberspace. The writer’s lasting cultural impact and prominent status within the CCP canon make him an apt weapon and a shield when navigating constant (and often erratic) censorship and erasure on platforms such as Weibo. While the censors do eventually catch on, a far more intelligent algorithm than those currently in use would be required for autodetection and prevention. It is an indirect yet nimble form of speaking back to power precisely by mobilising its sanctioned voices.

One radical implication of the ‘Lu Xun bot’ is the future potential of bringing cultural luminaries such as Lu Xun back to ‘life’ using an advanced AI system. Although legendary artists such as the late Hibari Misora have been revived in stunning performances enabled by VOCALOID:AI, the prospect of AI-enabled cultural criticism still feels distant, with its necropolitics, ethics, and ontology remaining rather murky. On the other hand, the ‘Lu Xun bot’ feigns a sense of machine-induced objectivity and randomness, hence escaping the liabilities and political consequences that would be activated by a concession of human agency. Self-deprecating yet defiant, this satirical gesture belies a sense of

resignation that speaks volumes about the actual state of human agency, which seeks anonymity and shelter through nonhuman camouflage in a mass surveillance state. As Chinese netizens make creative censorship-circumvention a national pastime—one may argue this is one of *the* cutting edges of cultural production—the swift and constant erasure of these strategies, data points, and new semantics make scaled, deep learning unfeasible; let alone history-writing of any feasible continuity or validity. One wonders not only how that zeitgeist might be captured or modelled, but indeed how it might also be transformed as people intuitively adapt to coded speech in Chinese cyberspace.

What is profoundly odd and ironic here is a palpable sense of machine envy: advanced technology doesn't necessarily embody the all too familiar tropes of servitude or existential threat, but rather presents a viable, aspirational *model of how to be*. Artificiality not only feels more desirable but also more tangible than reality; often artificiality *is* the fabric of reality, particularly in the post-Socialist state. In his book *From Newspeak to Cyberspeak: A History of Soviet Cybernetics*, science historian Slava Gerovitch recounts a prescient insight from cultural theorist Mikhail Epstein that 'Soviet Marxist ideological language is by its very nature artificial; it would be possible to outline its structure using abstract formulas', and that 'with specific formulas of functions and markers, [a] computer would be capable of composing Soviet ideological texts'.[3] Epstein's words would find uncanny resonance in contemporary times, suggesting that a machine is indeed better suited for certain genres of statecraft.

On April 19, 2022, a sound art 'installation' created by Shanghai-based artist Yang Xiao went viral on Chinese social media. Entitled 'Goodbye to Language', a borrowing from Jean-Luc Godard's filmic inquiry into the enigmatic nature of communication, the video opens with views from the artist's apartment window onto a narrow alley in Shanghai during the indefinite, draconian quarantine of April 2022; the streets are empty except for those rushing about in white PPE coveralls. A loudspeaker blasts gibberish sentences in a standard newscaster voice: 'focus on the institution and schema, follow the law of the grand scope, expand the maximum limit, gradually improve interaction…'. This gibberish, however, is barely distinguishable from the newspeak that characterised much state media output of the time; the style is also informed by public announcements

intended to boost morale through the ideologically charged ‘pep talks’ with which citizens tend to be inundated during times of national emergency ranging from floods to earthquakes to pandemics. ‘Goodbye to Language’ was created by algorithmically randomising some six hundred words frequently used in state media, and voicing them with an AI newscaster found in iFlytek text-reading software.[4] Although destined to be repeatedly erased, with the discussions surrounding it hushed, the work self-actualises precisely through its mechanism of making and circulation—and ultimately in the widespread, unspoken understanding of the inarticulable absurdity of the system of meaning/nonsense that the work so powerfully reflects back to itself. Both its virality and its erasure demonstrate a profound, collective understanding of the infrastructure of power and language governing everyday life. It is also not surprising that when ChatGPT, OpenAI’s wunderkind chatbot, caught on, netizens immediately thought of using it to generate personal statements in applications to join the Chinese Communist Party, which typically demand a highly schematic structure and tone. ChatGPT’s results were predictably astounding.[5]

An understanding of this distinct type of newspeak—opaque, ideologically suffused, and often inherently contradictory—also explains the appeal of the ‘Soviet Jokes’ genre, which enjoys a robust presence in contemporary Chinese discourse: from accounts like the aforementioned ‘Soviet Jokes bot’ which dispenses comically timed political anecdotes, to being used as an adjective for describing singularly absurd realities, to the ‘Soviet Jokes Generator’ created in 2021 by GitHub user ‘disksing’,[6] a database developer based in Beijing.[7] Soviet Jokes Generator requires input in six simple categories: the subject of satire; the entity that raised the issue; the subject’s alleged benefits according to the entity; the victim; the organisation. The landing page provides some curious default entries in the aforementioned categories: respectively, voting to implement a ’996 workweek’ (the controversial scheme of 9 am to 9 pm workdays, six days per week); the HR department of Pinduoduo (an extremely popular e-commerce platform specialising in low prices and bulk purchases); auspicious returns; workers; Pinduoduo. One possible outcome reads: ‘Pinduoduo HR surveys a worker’s opinion on the implementation of the work week, to which the worker replied: “I have different options, but I object to them!”’.

Default page of *Soviet Jokes Generator.*

The fact that the Soviet Jokes Generator uses the Chinese language and speaks to Chinese realities without evoking any explicit Soviet reference at all is significant. Decades after the official Sino-Soviet split, the Soviet aesthetic-political project still exerts a nuanced influence on China's (post-) modernity, which manifests in art pedagogy, monuments, familiarity with Russian/Soviet literature and music, urban planning, newspeak and, say, Ilya Repin's paintings, which have been immortalised as illustrations in national textbooks with an estimated readership of a hundred million. Artist and theorist Hito Steyerl has recently explored the notion of 'artificial stupidity', both in the sense that we overestimate the power of advanced technologies, and that the low-grade deployment of artificial stupidity can easily lead to dire consequences, such as Twitter armies that effectively sway politics but require only simple coding. AI discourses concerning China have seemingly been dominated by such totalising narratives around geopolitics (e.g. AI races between China and the US), panopticon-like cybernetics and mass surveillance, and the accelerated embrace of new technologies both on the state level and in the cultural sphere. Yet understanding the cultural ramifications of AI is a more intricate task. What kind of cultural understanding can AI harness, produce, and illuminate, particularly against extensive censorship, swift evolutions in coded speech and neologisms, and insufficient understandings (and biases) of historical mediation and its cultural affinities? In short, what is post-Socialist AI?

Dislocation (1986), dir. Huang Jianxin.

Allegories of Bodies

Before clones, Siri, Sophia, Tay, and virtual influencers, director Huang Jianxin created a distinct prototype—the AI communist bureaucrat—in his 1986 film *Dislocation*.[8] Inspired by the utter fatigue of obligatory and meaningless meetings, the film finds its protagonist, a robotics engineer-turned-manager, creating a robot in his image to handle meetings and other bureaucratic nonsense, so that he, a true scientist with the temperament of an intellectual, may focus on research and invention. Performative bureaucracy isn't exactly a common task for machines, yet it is essential to culturally specific workplace politics; as the protagonist ponders: 'If I don't comply with the tasks I detest, I don't get to do the work I like.' Over time, however, the AI grows curious about the world outside its assigned activities and becomes eager to explore romantic love and learn how to hold a cigarette elegantly; it also genuinely enjoys being a powerful bureaucrat. Eventually, the AI threatens to replace the scientist, who is forced to terminate his experiment.

The film unfolds in a surreal, symbolism-laden Bauhaus setting—a rather bold and refreshing aesthetic choice for Chinese cinema in the 1980s. It also channels an unnerving atmosphere of perpetuity that often characterises corporate and totalitarian environments and which has resurfaced in more recent TV shows (such as *Severance*, 2022) that explore workplace dystopias. While the plot of the intelligent machine gone awry might now feel familiar, even cliched, the AI bureaucrat in *Dislocation* was already philosophising about the deep existential ironies that David Graeber would later analyse through the notion of the 'bullshit job', forms of salaried employment that are utterly pointless and that exist 'just for the sake of keeping us all working'.[9] In *Dislocation*, bureaucratic 'bullshit' work maintains the ideological structures unique to Chinese workplaces (even to

this day). At the same time, the AI bureaucrat embodies existential anxieties unique to a China in the throes of modernity—constantly besieged by new information and new longings in a present fragmented by multiple histories and multiple futures.

In both appearance and temperament, *Dislocation*'s AI bureaucrat presents a profoundly human machine, capable of joy and philosophical thought, which approximates neither earlier Socialist prototypes of heroic and altruistic model workers, nor the neoliberal individual whose agency is reduced to statistics under digital governance. These prototypes and allegories of the technologised body merit further analysis, as they inform cultural thinking around bodily relationships to machines and algorithms. In her book *The Fate of the New Man: Representing and Reconstructing Masculinity in Soviet Visual Culture, 1945–1965*, historian Claire E. McCallum explores how Soviet ideology crucially links work and masculinity, glorifying productive labour. Citing Maurizio Boscagli's observation that 'the modernist agenda was not simply to emulate and challenge technology but instead to identify with it', McCallum notes the resulting shift in the aesthetic appreciation of the male body, which emphasises its productive capacity and use.[10] The ripple effect was felt throughout the Second and the Third World. In the 1960s, Chinese oil driller Wang Jinxi would enjoy national renown as the 'Iron Man' for his astounding devotion and productivity while spearheading the development of the Daqing oil fields. Another household name of the era, Lei Feng, propagated the 'screw nail spirit', which highlights the importance of collective striving and the nobility of fulfilling humble yet essential roles in the greater Socialist project.

Left: Liu Qing, *Youth Aflame* (1974), a painting created in the Socialist Realism idiom celebrating a female worker's prowess in operating a blast furnace from the control room. Right: An ethnic Tibetan woman in China's Sichuan Province carrying components of decommissioned cryptocurrency mining machines, 2021; in its composition, the image is strikingly redolent of another genre of Socialist Realism valorising China's ethnic minority groups engaging in productive labour (Photo by Ding Gang, Caixin).

The 'screw nail' allegory certainly implies the scale of collectivisation, too, which finds resonance both in China's demographic dividend and in the tragic exploitation of, for instance, delivery persons and cyber content analysts under the algorithmic optimisation of performance metrics. The spectre of the 'human flesh search engine'—*renrou yinqing* (人肉引擎), where the searching capacities of random netizens are powerfully mobilised at scale, looms large in both cyberbullying incidents and social justice efforts, particularly when the omnipresent surveillance grid (curiously) fails —as in the recent case of a still unidentified woman chained at the neck by her husband in a dark chamber in rural Xuzhou. Glimpses of the woman's abhorrent living conditions were first shared by a TikTokker who visited the impoverished family to bring supplies; national outrage and netizen journalism quickly ensued. The cognitive dissonance came not only from the egregious violation of her bodily autonomy (chained and forced to have eight children) in a relatively developed area of the country officially purporting to adhere to Socialist feminist ideology, but also from the sheer

difficulty of tracking down her identity and family information within a state known for advanced surveillance technology and deployment—so advanced, sometimes, that people have had their Covid passes pre-emptively blocked because records exist of their shopping for cold medicine. It appears that the ultimate obstacle course for human agency lies between the extreme sensitivity and the extreme negligence of 'smart' governance.

At the same time, AI-related vernacular often sheds light on the dynamic evolution of cultural attitudes. The term 'AI *yanji*' (AI演技) or 'AI acting skills' is widely applied to describe actors who lack emotional and expressive range, while the Japanese neologism 'Chi-borg' describes Chinese beauties with impeccable makeup and looks, which certainly belies both cultures' (somewhat problematic) assumptions about female beauty. 'Under the skin'—'*pixia*' (皮下) refers to the actual personalities behind internet personas, whereas '*zhongzhiren*' (中之人) or 'middle-human' denotes performers of fictional roles, which can encompass both physical mascots such as the Disney Shanghai phenomenon LinaBelle, and virtual idols such as A-Soul—a popular VTuber girl group created by ByteDance and Yuehua Entertainment. Human actors are often credited for giving virtual idols distinct personalities, quirks, and affective engagement with their fanbase, but overwork and other forms of exploitation are obscured by the 24/7 veneer of virtual beings and the invisibility of the actors' labour. 'Tool people', '*gongjuren*' (工具人) refers to those who willingly put themselves at others' disposal in various relationship dynamics, usually to the detriment of their own financial and mental health, and without expectation of reciprocal affection.

While many of these terms have neutral or humorous connotations, there are also more egregious examples, such as 'lining'—'*neidan*' (内胆) originally meaning the inner lining for clothing, but also used as a derogatory reference to human performers in heavy fursuits, such as the human 'mascots' in the shopping malls and theme parks of the world. Science fiction writer Ted Chiang has long warned that before AI can become intelligent enough to harm humans, it will first and necessarily experience pain and cruelty, citing abundant evidence of the abuse of animals and small children, and certainly of other humans, under capitalism.[11] The moral implications of terms like 'lining', which crudely and literally equate humans with objects, testify to Chiang's insight, as do

the terse orders from the scientist in *Dislocation* to his AI bureaucrat: ‘I did not create you so that you can express ideas, but to obey my directives; otherwise I’d rather destroy you.’

Souls, or The Time to Live and the Time to Die

Questions regarding the soul—or ensoulment—in relation to AI seem to focus heavily on ontological metrics and the computational possibilities of AI. Literature, art, and videogames of late, however, have meaningfully relocated this discourse to soulful experiences with or in relation to AI, asking not what *it is*, but *what it does* and means (to us). *The Lifecycle of Software Objects*, Ted Chiang’s 2010 novel, is presciently insightful on the topic of human-AI relationships. In the story, rudimentary AI entities known as ‘digients’ are groomed as virtual pets, yet their survival is precariously dependent upon market demand and technological obsolescence. When the company goes bankrupt, the love and care of their trainers (one of them fittingly a former zookeeper) keeps the digients ‘alive’, and, more importantly, makes their growth meaningful.[12] Real-life counterparts of the story can be found in the Buddhist funerals held for decommissioned AIBO robot dogs when Sony announced in 2014 that they would no longer support upgrades and maintenance for the aging robots. During the product’s timeline from 1999 to 2004, many AIBO owners had grown deeply attached to their AI pets and considered them family members whose passing merited a deeply spiritual reckoning and remembrance.[13]

There is a parallel between deep care for software/virtual beings and the vast possibilities opened up by digital realms housing errant emotions. When Nintendo’s *Animal Crossing*[14] became an explosive success during the COVID-19 pandemic, many users marvelled at the healing properties of the game’s design: synchronicity with natural time and seasonal creatures, worldbuilding through basic tasks and bartering, abundant and detailed cultural references. The limitations of the island’s visitors—anthropomorphised animals—moulded players’ behaviour. Some gamers would intentionally ignore or harass animals they disliked in order to make room for more popular animals, with many becoming dismayed at the realisation of their own cruelty, sharing feelings of shame on social media when the unwanted animals showed grace and gratitude as they packed up

to leave. The inter-user features of *Animal Crossing* have also expanded human relationships, overcoming not only geographical distance but also the boundaries of life and death, with deceased loved ones leaving behind virtual assets as gifts, or entire built and lived-in environments for their friends and families to visit.[15] Over time, the necropolitics of both virtual assets and 'software objects' will certainly transform our understanding of the material culture of death, and by extension our sense of memory, community, and history.

But not all digital resurrections are equally desired. The existential anxiety of not fully erasing one's social media records before death has become widely memeified; after all, if all our transactions, border crossings, criminal records, medical bills, and job histories are automatically archived and relentlessly harvested as big data, we can never truly cease to be. At the same time, being with the digital avatar of a loved one who has passed away is drastically different from seeing their photographs newly animated, a function many AI tools readily provide. The punctum, Roland Barthes's notion of the specific thing in a photographic image that uniquely pricks us ('but also bruises [us], is poignant to [us]'), is both magnified and perverted in old photos reconstructed and animated by AI.[16] If photographs are evidence of 'that which was there', these AI-reanimated images displace nostalgia to a temporally confused condition of 'what could have been'.[17]

Posted by u/MajesticSamm 6 months ago
40.1k
My best friend passed away in May but thanks to her sister setting up the dream address, I got to see her this morning.
New Horizons

Above: A reddit user sharing the experience of ‘reuniting’ with their deceased best friend in *Animal Crossing*, 1. Below: Still from the NHK Special *Bringing Hibari Misora Back with AI*, 2019.

As touched on above, in 2019, Yamaha Corporation revived the legendary musician Hibari Misora in a poignant performance using VOCALOID:AI through a collaboration with NHK’s annual kohaku special program, which marked the first real world implementation of the singing synthesis technology. Using actual recordings of the late artist’s songs and speech as machine learning data to render a new score, VOCALOID:AI can generate the ‘unique mannerisms and nuances of the vocalist with any melodies and lyrics’.[18] In the live performance, the cultural icon appears as an animated

hologram, and the song, entitled ‘Are Kara [*Ever Since Then*]’, poetically evokes the passage of time: ‘The horizon hasn’t changed, but the landscape quietly transforms / to live is to understand the meaning of parting.’ The broadcast also captured many audience members wiping away tears, especially when the singer directly addressed her audience in the lyrics ‘My beloved, how have you been since that day?’—a line that also gives one chills for its acknowledgement of the AI singer’s agency as Hibari Misora’s ‘could have been’. As deeply moving as the performance was, however, it also sparked debates about the right to be forgotten, or to be remembered rather than ‘revived’—a branch of afterlife management that will always tread the fine ethical line between transgression and veneration.

On Art and Cultural Specificity

During the COVID-19 pandemic, I served on the jury panel for several zoom-based critiques, where Chinese and sinophone artists across time zones presented new works and workshop ideas. An overwhelming majority of these emerging artists deployed GANs: generative adversarial networks, where two neural networks are pitted against one another to generate new, synthetic data that can pass for real data. In the meantime, being based in New York, I witnessed the exhilarating rise of Asian American artists, many of which address racialised trauma, the Orientalist gaze, colonial/diasporic history, and issues of identity and representation in their work. This distinctly American discourse of identity politics which powerfully informs biopolitics and art making usually holds little sway in the cultural context of China. In both places, observable trends often reflect not just artists’ intentions, but also the kinds of cultural mediation that render certain topics more visible, popular, or accessible. When cultural markers do appear in works that engage with AI, they are often telling of underlying cultural assumptions and creative strategies.

Above: Ian Cheng, *Life After BOB: The Chalice Study* (2021), live animation, 48 min. Courtesy of the artist. Below: Miao Ying, stills showing one character from *Pilgrimage to Walden XII: The Honor of Shepherds* (2019–2020), a six-screen game engine live simulation with machine learning software run on computers or mobile devices.

The Chinese American artist Ian Cheng has been a long-time pioneer of algorithm-based art and open-ended simulations. His film *Life After BOB*[19] (2021), a narrative animation built in the Unity game engine and presented live in real time, speculates on the feasibility of an AI-optimised life with chilling, psychological realism. In the artwork, Dr Wong, a neural engineer, implants an AI system into the brain of his ten-year-old daughter Chalice to guide her through life, hypothetically testing and ruling out errant chains of action to fulfil Chalice's destiny. A profound dilemma lies at the heart of the algorithm's success: a life well lived—ensured by the AI's guidance—is lived without real participation; Chalice is completely alienated by her autopiloted alter-ego. Parenting emerges as a prominent subplot. Is Dr

Wong's experiment a daring revolution or an extreme form of helicopter parenting? This dilemma of parental guidance and care finds curious resonance in the notion of 'lifting' in *Kazuo Ishiguro's Klara and The Sun*, where children may go through an elective procedure that significantly elevates their intelligence, ensuring success in adult life; but 'lifting' might also kill them prematurely.[20] It is not hard to recognise wider echoes of the education race and extreme 'involution'—*neijuan* (内卷)—in Asian and Asian diasporic communities.

Created around the same time and similarly philosophising 'desirable outcomes', New York-based Chinese artist Miao Ying's generative film *Pilgrimage to Walden XII, Chapter I: The Honor of Shepherds*[21] encodes its cultural references far more obliquely. Set in a reimagined mediaeval European village, a popular setup for many fantasy games, *Walden XII* is governed by a digital indulgence system entitled to 'explain its ideological doctrine to its people, implement technologies that enforce good citizen behaviour, collect behavioural big data, and reward or punish citizens according to behavioural scores'. The work itself showcases six AI 'Social Shepherds', each representing a different social class and level of intelligence, visualised as distinct characters in varying landscapes. A dynamic score is displayed across or adjacent to each 'Shepherd', reflecting their competency at the task of 'model citizenry'. Spelled in Leet, the characters' names provide thinly veiled clues: Brain Urban Migrant W0rk3R, Brain Outsourced Pr0l3tar1at, Brain B0urg3o1s, or Brain W01f W4rr10r, the latter borrowed from the blockbuster franchise Wolf Warrior,[22] which recentres the Hollywood lone-hero narrative on a retired, deeply patriotic Chinese soldier.

While *Life after BOB* features fantastical reimaginings of cultural signifiers such as the reclining Buddha, Miao Ying deliberately eschews visible cultural references or direct commentary on the work's meaning. Both in its title and concept, however, *Walden XII* references *Walden Two*, the utopian novel written by behavioural psychologist 1B.F. Skinner on yielding desired behaviours through reinforced positive conditioning.[23] The artist's preference for codified meaning clearly resonates with broader tendencies among Chinese netizens, such as those discussed above, although it also functions as a buffer, lest the work be immediately reduced to a mere satire on censorship and surveillance. The artist, whose practice closely charts the history of Chinese cyberspace, finds her creative locus in

questioning both open and closed systems, often favouring the more ‘stupid’ algorithms. After all, culture tends to manifest most potently in glitches, mistakes, absurdities, latent assumptions, and transient nuance. One need look no further than the limited, tired repertoires of, say, Vermeer, Magritte, and Van Gogh in AI-related visual experiments to recognise the already ubiquitous biases in the data, which means that we certainly need to look harder still—not only at vastly different datasets, but also at the assumptions that govern them and the parameters that instrumentalise them. As illuminated in the covert strategies through which Miao Ying works with artificial intelligence, machines might not deliver economic or political emancipation, but they might generate important kinds of agency through camouflage, digression, divergence, and misalignment.

Notes

1. E. Huang, ‘A bot tweeting passages from modern China’s most famous writer has been silenced’, *Quartz*, 24 July 2019, <https://qz.com/1673653/china-shut-down-a-bot-tweeting-words-from-famous-writer-lu-xun/>.
2. Lu Xun (鲁迅), ‘Miscellaneous Thoughts’ (小杂感) (1927), <https://www.marxists.org/chinese/reference-books/luxun/11/030.htm>.
3. S. Gerovitch, *From Newspeak to Cyberspeak: A History of Soviet Cybernetics* (Cambridge, MA: MIT Press, 2002), 3.
4. Xie Ding (谢丁), ‘Goodbye Shanghai, Goodbye Language’ (再见上海，再见语), *Career Engine*, 9 May 2022, <https://posts.careerengine.us/p/62791d001f18d5382dee5b35>.
5. An example can be found on the social media platform RedNote (小红书, more literally ‘Little Red Book ’): <https://www.xiaohongshu.com/explore/6394a2ac000000001c00516a>.
6. The username seems to reference ByteDance, the Chinese multinational internet technology company that owns, among many things, TikTok.
7. See <https://disksing.com/sao-gen-gen/3> and <https://github.com/disksing/sao-gen-gen/issues/3>.
8. *Dislocation* (错位), dir. Huang Jianxin (黄建新), 1986.
9. D. Graeber, ‘On the Phenomenon of Bullshit Jobs: A Work Rant’, *STRIKE! Magazine*, August 2013,

<https://web.archive.org/web/20210730130503/https://www.strike.coop/bullshit-jobs>.
10. C.E. McCallum, *The Fate of the New Man: Representing and Reconstructing Masculinity in Soviet Visual Culture, 5–1965* (Dekalb, IL: NIU Press, 2018), 6.
11. E. Klein, 'The Author Behind "Arrival" Doesn't Fear AI. "Look at How We Treat Animals."', *The New York Times*, 30 March 2021, <https://www.nytimes.com/2021/03/30/opinion/ezra-klein-podcast-ted-chiang.html>.
12. T. Chiang, *The Lifecycle of Software Objects* (Burton, MI: Subterranean Press, 2010).
13. J. Burch, 'In Japan, a Buddhist Funeral Service for Robot Dogs', *National Geographic*, 25 May 2018, <https://www.nationalgeographic.com/travel/article/in-japan--a-buddhist-funeral-service-for-robot-dogs>.
14. *Animal Crossing*, Nintendo, 2001–2020.
15. See for instance the following post by u/MajesticSamm, *Reddit*, 1 December 2021, <https://www.reddit.com/r/AnimalCrossing/comments/r6ejl9/my_best_friend_passed_away_in_may_but_thanks_to/>.
16. R. Barthes, *Camera Lucida*, tr. R. Howard (London: Vintage Classics, 1993), 27.
17. Perhaps this is why I could not bring myself to reanimate my late mother's photographs.
18. 'Yamaha VOCALOID:AI™ Faithfully Reproduces Singing of Legendary Japanese Vocalist Hibari Misora', *Yamaha*, 8 October 2018, <https://www.yamaha.com/en/news_release/2019/19100801/>.
19. *Life After BOB*, dir. I Cheng, interactive animated series, 2021.
20. K. Ishiguro, *Klara and The Sun* (London: Faber & Faber, 2022).
21. Miao Ying, *Pilgrimage to Walden XII, Chapter 1: The Honor of Shepherds*, installation with live simulation, 2019–2020.
22. *Wolf Warrior* (战狼), dir. Wu Jing (吴京), 2015.
23. B.F. Skinner, *Walden Two* (Indianapolis: Hackett, 2005).

Blaise Agüera y Arcas

The Mother of Ten Thousand Things

In the silence and the void,
Standing alone and unchanging
Ever present and in motion.
Perhaps it is the mother of ten thousand things.
I do not know its name.
Call it the Way.

LAOZI[1]

Ten Thousand Coins

Over the last decade, LA-based artist David Lebrun has been making stunning short videos of traditional art and artefacts. These range from palaeolithic stone tools to Karnatic temples, to Irish standing crosses, to medieval Iranian finials. The videos are not only beautiful but also offer, in the way art sometimes can, a new way of seeing the world and ourselves.

Lebrun's technique begins with dozens or even hundreds of still images of a specific class of artefact—for example, in his work *137 Coins*, gold staters, with both faces of the coin shown side by side against a black background. These shots are arranged into a sequence based on visual similarity, or chronology, which often amounts to much the same thing. A Greek stater in the fourth century BCE, for instance, features a realistic rendition of Philip of Macedon on the obverse and a two-horse chariot on the reverse. Three centuries and one hundred and thirty-seven coins later, both human and horse have undergone such transformation and abstraction that at times it is impossible to tell what the designs represent. Although many conscious aesthetic choices were involved, it seems likely that some of the artisans who engraved these designs were equally in the dark. Certainly, none of them could step back to see the whole. Like neurons in a giant brain or filing clerks in the great Chinese Room of human culture, they were vehicles for the creation and transmission of a numismatic 'text' that transcended their own understanding.

David Lebrun, *137 Coins, Greece via Rome to Gaul (4th to 1st C. BCE)* (2020), DCP, colour, 7 min. Stills showing an early, realistic rendering of a Greek stater juxtaposed with a later, far more abstract Gallic rendition.

Because stylistic evolution happens by degrees, no two consecutive artefacts look that different. This makes it possible for Lebrun's video to morph convincingly from one image to the next. If it weren't for a momentary pause at each original shot, it would be difficult to tell which are the interpolated 'fictitious' artefacts, and which are 'real'. At about seven minutes long and twenty-nine frames per second, *137 Coins* contains roughly ten thousand frames, only about one percent of which are images of actually-existing coins. The rest of these ten thousand things, like deepfakes, merely could have existed—and perhaps do exist, in some parallel universe. They belong to the 'adjacent possible'.[2]

Or perhaps this is the wrong way to see things. Here is an alternative. *137 Coins* is a form of time-lapse photography rendering certain kinds of actually-existing objects visible for the first time, just as false colour images

made using X-ray instruments let us see what we cannot with the naked eye. X-ray telescopes, for instance, allow us to see the black hole at the centre of the Milky Way, and X-ray crystallography allows us to resolve the structures of complex molecules. To see such things, photons in the visible range, transduced into electrical impulses and then into meaning by our unaided eyes and brains, are not sufficient. Computer-assisted transpositions in time, scale, and frequency are required, often along with some further image processing. In the same spirit, Lebrun's short film compresses time by a factor of about twenty-six million, combining sparse source material with the computational magic of temporal superresolution. If we cannot perceive them directly, are black holes and proteins real? What do we mean when we assert that they are? And what are we to make, then, of the centuries-long cultural object (or *hyperobject*)[3] revealed by *137 Coins*? It is, I would argue, every bit as real as a black hole or a protein.[4]

The same holds even when we return to the familiar mesoscale world. Consider what happens in our brains when we render a jerky manifold of visual impressions into a 'butterfly'. Every such act of perception involves a semantic abstraction. Our brains have evolved to be good at these abstractions because they are helpful for predicting the future: if we could become micro-cosmonauts, 'see' a protein at a distance and identify it, we could predict which receptors it will likely dock with, just as identifying a butterfly in flight lets us make guesses about its future trajectory, and its likelihood of landing on a flower. This is also why self-supervised neural nets trained to predict token sequences learn concepts. The salient point here, though, is that the imponderable question 'Is it real?' can be restated information-theoretically as 'Is it a helpful abstraction for recognising an anomaly or predicting the future?'. When we make a perceptual error, for instance believing that we saw a cat out of the corner of our eye, then turning to see it was only a scarf draped over a chair, it is an error of prediction, which we experience as surprise.

Let us now turn from semantics to style. Semantics is *what* we see, which *137 Coins* holds constant throughout its ten thousand frames. Style is *how* we see, or how what we see is rendered, which varies dramatically over the three centuries documented by Lebrun's work.

Ways of Seeing

As German museum curator Max Friedlander noted in *Von Kunst und Kennerschaft*, ‘Art being a thing of the mind, it follows that any scientific study of art will be psychology. It may be other things as well, but psychology it will always be’.[5] Friedlander’s use of the term *psychology* here was old-school, more William James than Sigmund Freud. He meant neuroscience.

Today, the neuroscience of vision, which especially concerned Friedlander, is among the best-studied from a holistic, computational, and functional perspective. It also marks the point of closest approach between computational neuroscience and AI. In particular, most major advances in AI throughout the 2010s were based on convolutional neural nets (CNNs), whose architecture mirrors the classic model of visual perception developed in the second half of the twentieth century by David Hubel and Torsten Wiesel (winners of the 1981 Nobel Prize in physiology and medicine).

By conducting experiments in which visual cortical neurons of anaesthetised cats were electrically recorded in response to projected images, Hubel and Wiesel learned that vision is organised hierarchically. Cells early in the visual pathway, beginning in the retina itself, are sensitive to highly localised features—if not exactly ‘pixels’, at least small spatiotemporal blips. As one ascends the neural hierarchy, from the retina to the lateral geniculate nucleus (LGN) to the primary visual cortex (V1) and beyond, neural responses become less localised and more semantic. During development, we learn to see by modifying our neural connectivity, forming, strengthening, and weakening synapses between neurons at one stage and the next so that the network as a whole can perform the key function of vision: recognising stable, meaningful things out there in the world. Early in the visual cortex (during the first stages of processing), neurons respond to local edges, while by V4 they respond to shapes. Subsequent researchers extended Hubel and Wiesel’s work by correlating neural output with visual noise stimuli (a bit like a static-filled TV screen), allowing them to build computational models of the ‘receptive fields’ of these neurons,[6] from ‘pixels’ and ‘edges’ up to entire objects and people. In 2005, a Jennifer Aniston neuron was even discovered while doing exploratory recording from the brain of a human patient.[7]

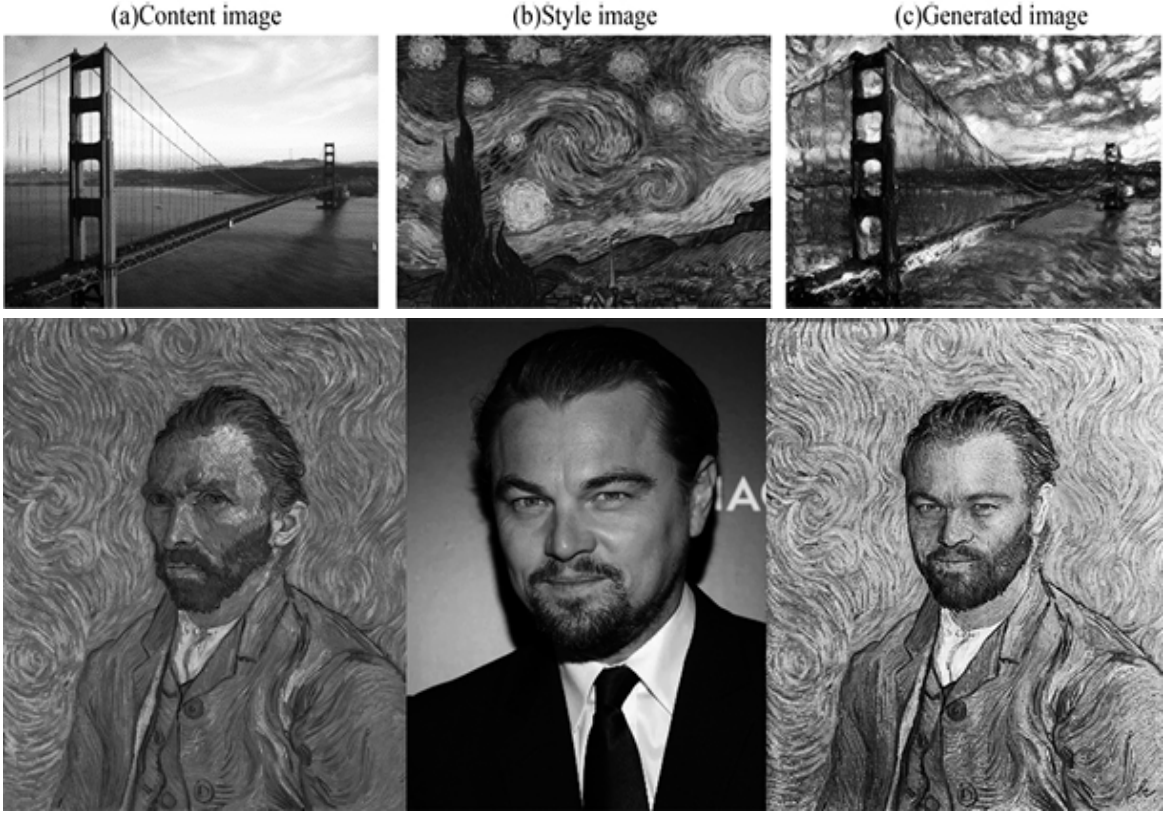
The great triumph of CNNs was that not only did they work better than any prior approach to machine vision in practice, they also bore an uncanny functional resemblance to the visual cortex.[8] ‘Deep’ convolutional networks

were originally designed to classify images. Their first layer consists of the input pixels (analogous to the photoreceptor cells in the retina), with each subsequent layer taking input from a localised region of the previous layer. Merely by wiring simplified artificial neurons into a Hubel-Wiesel style hierarchy and training the network to classify objects, researchers saw ‘edge neurons’ emerge in earlier layers, and increasingly complex, non-localised forms emerge in later layers, just as in the brain.[9] Artificial neurons in the final layer of a CNN are, by construction, trained to respond uniquely and specifically to high level semantics, such as (in the case of a network trained to recognise celebrities) Jennifer Aniston.

The internet phenomenon known as ‘DeepDream’ started when Alex Mordvintsev, a researcher at Google, sought to create better visualisations of the receptive fields of artificial neurons in CNNs after they had been trained.[10] During the training process, the weights (or, in biological terms, the synapse strengths) of the connections between the neurons in a net are optimised to create a working visual system. Mordvintsev instead fixed the weights of a trained network but optimised the image itself to produce maximal activation in a given (artificial) neuron. He began with pure noise images, and watched patterns emerge—either simple, if the neuron was early in the network, or complex and semantic, if the neuron was near the output layer. This was reminiscent of the visual noise techniques neuroscientists had used to estimate the receptive fields of real neurons, and the results were similar. When Mordvintsev tried starting with natural images rather than noise and optimised the pixels to enhance the overall activity of neurons in some middle layer of the network rather than specifying an individual neuron, the resulting images could be beautiful, hallucinogenic, and sometimes frightening. They reminded many of LSD experiences. Some neuroscientists believe this is no coincidence; psychedelics may enhance the role of expectation (or ‘top-down feedback’) in ways that produce a similar result.[11]

In the summer of 2015, around the same time as DeepDream, neuroscience postdoc Leon Gatys and collaborators in the Bethge lab at the University of Tübingen invented a related technique called ‘neural style transfer’.[12] The motivating insight was that two images of, say, a fish, one rendered in the Ukiyo-e style and another in an the style of a European impressionist painter, should activate the same ‘fish’ neuron in the final layer of a CNN; however, clearly, the pixel values are very different, and

presumably, the patterns of activation in the middle layers of the network will be different, too. Yet if a style is recognisable in its own right, independent of the image content, there must be something about the overall pattern or statistics of these middle layer activations that say ‘impressionist’ or ‘Ukiyo-e’. It should then be possible to take image A, and re-render it in the style of image B, by using optimisation to modify the pixels of A (as DeepDream did) to reproduce the middle-layer neuron activation statistics of image B while preserving the output layer activations.



Examples of style transfer from L. Gatys, A. Ecker, and M. Bethge, ‘A Neural Algorithm of Artistic Style’.

Gatys and coauthors made the simplest possible assumption about what might comprise the statistics we call ‘style’: pairwise correlation between neurons, meaning the degree to which every neuron in a middle layer co-varies with every other one. Astonishingly, it pretty much works. Neural style transfer lets one easily generate an image C by combining the content of image A with the style of image B. A flurry of websites and apps soon

sprang up to turn photos (and even movies) into van Gogh paintings, medieval tapestries, or nineties anime.
The basic premise of computational vision, as modelled by a CNN, is surprising enough: the semantic content of an image can be fully expressed in a not-so-large array of numbers representing the activations of the output layer. Visual style turns out to be expressible as an array of numbers too: the correlation matrix of middle-layer activations. In this sense, we now have a working functional model of the neuroscience, or as Friedlander would have it the ‘psychology’, of visual perception and representational style.

Ways of Rendering

Although partial and imperfect,[13] this model of style offers some profound insights; yet remarkably little has been written on this topic since 2015, perhaps because of the deep schism that has grown between the humanities and STEM subjects. The schism goes beyond differences in training or literacy, although that is a factor; it reflects deeper conflicts about human identity and values.[14] Let us set this wider sociological context aside, however, and consider how a neurally-informed approach to decomposing visual perception into ‘content’ and ‘style’ would allow us to explore important concepts in the history of visual art rigorously, and even quantitatively. Here are a few.

1. Individual Artistic Style

When an artist learns by imitating a master, whether through formal training or via exposure, the student’s style matrix will be similar, but likely not identical, to that of the master. When art historians or connoisseurs recognise an individual artist’s work, independent of the subject matter, it is this style matrix they have learned to recognise—encompassing fine details like brushwork or quality of line in the correlation structure of early layers of the neural net, and more holistic, thematic style at later layers.

2. Stylistic Mixing or Influence

It is straightforward to interpolate between style matrices, or equivalently, for an artist’s style to be influenced by multiple sources. By quantifying style, we can directly measure similarity between artists, and mark shifts in

similarity, as well as potentially inferring directedness of influence (who influenced who) over time.

3. Gradual Stylistic Evolution

Given imperfect transmission of style from master to student (hence, individual style), we would expect a gradual drift or evolution in style over time. If there is selection pressure (i.e. certain stylistic variations consistently find a better market or higher-status patronage than others) then this gradual evolution will be directed, or if undirected, we may simply observe ‘genetic drift’ in style space. Since selection pressure is subject to fashion, artistic evolution may be both directed and exhibit drift over longer timescales, resulting in successive stylistic regimes or ‘-isms’.

4. Realism

Some art historians, especially in the twentieth century, have insisted on total relativity in what comprises ‘realism’ in art. In his 1960 book *Art and Illusion*, Ernst Gombrich wryly noted,

> Perhaps [Rudolf Arnheim] is too eager to follow [Alois] Riegl in his ‘objectivity’, too eager also to vindicate the experiments of twentieth-century art to see the problem of illusion as anything but a Philistine prejudice. The fact that different periods are known to have had different standards of ‘lifelikeness’ makes him hope that a ‘further shift of the artistic reality level’ will make works of Picasso, Braque, or Klee ‘look exactly like the things they represent’. If he is right, the Sears Roebuck catalogue of the year 2000 will represent the mandolins, jugs, or twittering machines for sale on this new reality level.[15]

Clearly Gombrich was right to call BS. It *is* true that the realism of modern digital photography is far from ‘raw’ or ‘unprocessed’, but it is also true that natural images have a specific correlation structure or ‘style’, which can be more or less closely approximated by art—and which will thereby read as more or less realistic—even when such art remains semantically faithful to the natural image (i.e. it induces the same activations at the most abstract levels of the neural net).

There is an important ethological footnote here: ‘natural’ scenes in the city have a different style matrix from scenes in the forest, for example, and the differing statistical nature of the environments people inhabit doubtless colours their model of what is ‘realistic’, and indeed may even affect how they see. ‘The Dress’, an internet sensation that emerged the same year as

DeepDream and neural style transfer, illustrates this point. Photographed under ambiguous lighting conditions, the infamous Dress looks black and blue to some people, and gold and white to others. It turns out that the main determinant of which way the Dress is perceived appears to be whether the perceiver is an early riser (gold and white) or a night owl (black and blue); in this unusual case, differing 'visual priors' result not only in different style matrices, but in differing semantic perceptions.

5. Visual Metamerism

An influential 2011 neuroscience paper, 'Metamers of the Ventral Stream',[16] posed the following question: given some specific image A, and assuming a Hubel-Wiesel style hierarchical neural model of vision, what *other* images A' will be visually indistinguishable from A?

The researchers assumed that the viewer would fixate on a central point in the visual field. It is well-known that outside the fovea, our visual resolution is much lower, which might lead one to suppose that an unmodified image and one that is blurry outside the foveal region will be metamers, i.e. indistinguishable. Experimentally, however, this isn't the case. Peripherally blurry images *look* peripherally blurry. The researchers found instead that a metameric image A' could be constructed by perturbing the pixels of the original image A while keeping the 'average pairwise products' of V1 neuron activations—that is, their correlations—constant. This should sound familiar: it is precisely a case of keeping the style matrix (at the early, V1 layer) constant. The fixation point is important here because the sizes of V1 receptive fields increase away from the fovea, resulting in more and more obvious weirdness farther from the central fixation point—although, for a viewer fixating on that central point, this weirdness is, weirdly, invisible.

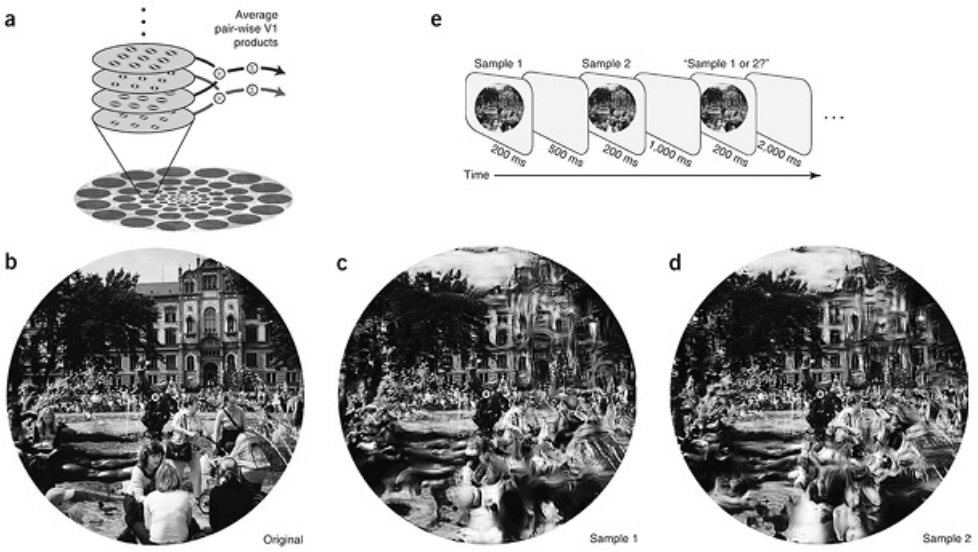
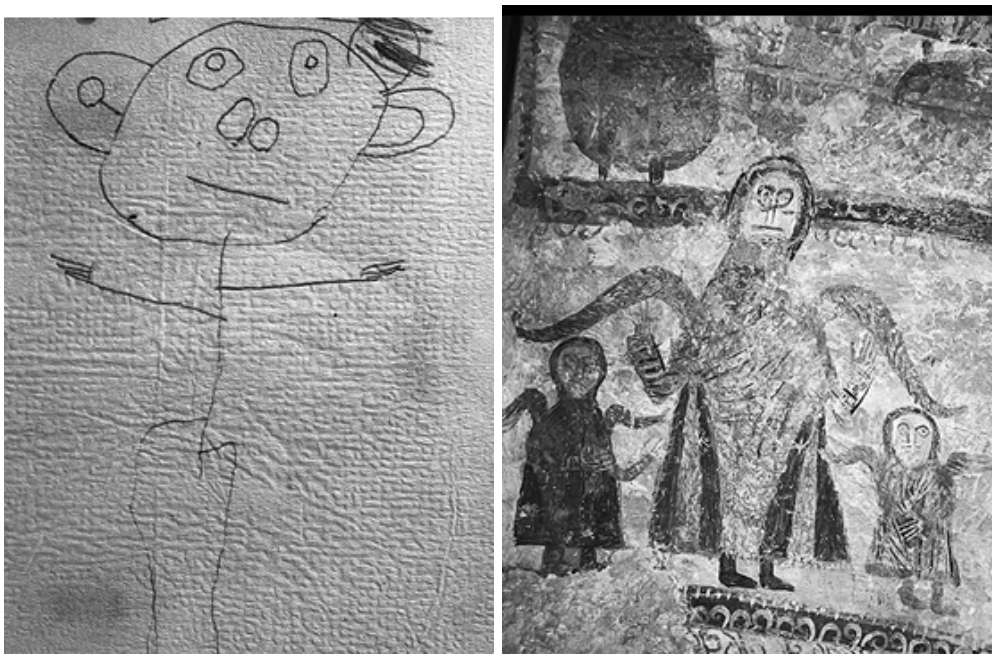


Figure from J. Freeman and E.P. Simoncelli, *'Metamers of the Ventral Stream'*.

What does this have to do with artistic style? For one, it predicts that if people's visual systems differ, either due to organic factors or to differing visual ecologies during development, their visual metamerism, hence personal style, will vary.[17]

6. Primitivism

'Primitivism' is a term even more fraught than 'realism', as it implies the supremacist practice (especially popular in the nineteenth-century Western anthropological tradition) of ranking cultures hierarchically from 'primitive' to 'advanced', with an accompanying value judgment. That we see the frequent embrace of 'primitive' style in cultures perfectly capable of 'realism', from Picasso's *bordel* (wherein self-conscious 'primitivism' allowed him to 'liberate an utterly original artistic style of compelling, even savage force')[18] to rage comix drawn in MS Paint, certainly runs counter to this notion. On the other hand, there is a clear connection between primitivism and rendering by young children or the unskilled, just as punk bands—the rage comix of music—embrace the aesthetics of unskilled musicianship.[19]



Left:A self-portrait by the author, aged 3, on a restaurant napkin.

Right: Detail of a Romanesque church fresco at the Museu Nacional d'Art de Catalunya, Barcelona.

What, then, does 'primitive' mean, in rigorous terms? One simple interpretation is that it is the statistical structure of 'default' rendering such that perceptual semantics are more or less preserved, but style correlations have not (or have not yet) been learned—either to copy the style of a master, *or* (in the case of visual art) to directly reproduce the correlational structure of a perceptual source. The exact nature of that 'default' depends on the sensorimotor system in question. For a generative model based on a CNN, the default is something like DeepDream, while for a human wielding a drawing tool, it looks more like children's art. The ability to generate art with 'default statistics' may or may not be preserved upon subsequent learning; if it is not preserved, like a language acquired in infancy but then forgotten, it must be explicitly re-learned.[20]

8. Abstraction and Non-Representational Art

Style transfer sheds light on visual abstraction, whereby visual elements and their relationships are selectively simplified while semantics are preserved. Abstraction is, put simply, just the name we give to a broad (and only approximately defined) region in style space. Taken to extremes, abstraction can begin to make it hard to figure out what an artwork depicts. At this

point, semantic content is no longer reliably conveyed, or may become decoupled from any original percept.

An art form can be representational to the extent that it reliably causes, in a third party experiencing it, the selective activation of high-level semantic neurons with meaning beyond the formal properties of the medium itself.

In music, this would seem to be less common, although it does occur—the technical term is ‘programmatic music’, meaning music that imitates natural sounds or relates to an extra-musical narrative. Such mimesis is obvious in shamanic traditions incorporating animal and bird calls, but one might assume most contemporary instrumental music to be non-representational, notwithstanding programmatic chestnuts like *Peter and the Wolf*, or more experimental mimetic works like John Luther Adams’s *The Wind in High Places*. Even in such cases, it is rare for a naive listener to hear the intended semantics without priming, although once primed, one cannot ‘unhear’ them. This is reminiscent of Mooney Faces, whose semantics may only snap into place with priming, but are then impossible to ‘unsee’.

A Mooney Face

But is there really such a thing as non-representational art? When one perceives an artwork, high-level semantic neurons *are* invariably and reliably activated, even if they are not ‘goose’, ‘wolf’, or ‘mountain peak’. Most naive listeners would describe *The Wind in High Places* as consisting of high whistling notes.[21] Perhaps the most we can say of so-called non-representational art is that it does not reliably activate certain *kinds* of semantic neurons, as in, for example, the traditional Muslim prohibition against visual mimesis of living things.

9. Style Collapse

A 2023 paper, ‘The Curse of Recursion: Training on Generated Data Makes Models Forget’,[22] describes the consequences of neural model inbreeding. The setup is straightforward: a first-generation large language model is trained on a large corpus of human-generated text. Then, this first-generation model is used to generate synthetic text, which is used as the training corpus for a second-generation large language model, and so on.

With every generation of this kind of recursion, the model deteriorates; the authors coin the memorable term ‘model collapse’ to describe the result. By the ninth generation, a language model prompted to expound on Perpendicular Gothic architecture will begin to gibber ‘In addition to being home to some of the world’s largest populations of black @-@ tailed jackrabbits, white @-@ tailed jackrabbits, blue @-@ tailed jackrabbits, red @-@ tailed jackrabbits, yellow […]’. A similar loss of structure and meaning takes place for other kinds of recursively trained AI models too, such as image generators.

Benjamin Bratton and I anticipated this problem (among others) in a *Noēma* piece published in July 2022:

> What happens when the number of AI-powered things that speak human-based language outnumbers actual humans? What if that ratio is […] 10:1? 100:1? 100,000:1? We call this the Machine Majority Language Problem. […] Nested within this is the Ouroboros Language Problem. What happens when language models are so pervasive that subsequent models are trained on language data that was largely produced by other models’ previous outputs? The snake eats its own tail, and a self-collapsing feedback effect ensues. The resulting models may be narrow, entropic or homogeneous; biases may become progressively amplified; or the outcome may be something altogether harder to anticipate.[23]

‘The Curse of Recursion’ offers experimental confirmation that, at least in a simplified lab setup, the Ouroboros Language Problem is real, and the resulting models are indeed ‘narrow, entropic or homogeneous’, with biases that become progressively amplified, i.e. per ‘The Curse of Recursion’, the ‘tails of the original content distribution disappear’.[24]

However, this result should not be interpreted simply as ‘what happens when AIs become decoupled from humans’. Rather, the authors are simulating what happens to cultural transmission over successive generations with zero recombinant diversity, or equivalently, with a population size of one per generation. We have seen similar ‘knowledge collapse’ or ‘complexity collapse’ in human populations, both under artificial lab conditions[25] and in natural experiments. The natural experiments have been described as ‘islanding’: human populations can

lose preexisting knowledge and technology if their numbers become too small, as has happened in islands with isolated populations like Tasmania[26] or, more dramatically, North Sentinel Island.[27] Over many generations, islanding can also begin to degrade the *actual* gene pool of a small population, resulting in losses of function (as in the famous 'island of the colour-blind')[28] or characteristic genetic disorders. We should expect such effects to be more rapid for memes (that is, cultural 'genes'), given the far greater speed of cultural transmission, mutation, and recombination.

Islanding has an effect both on semantic concepts and on style. Hence the primitive (in its technical sense) state of Romanesque frescoes like the one shown earlier is undoubtedly connected to the sharp decline in urban populations in Europe following the collapse of the Western Roman empire. The key to avoiding or reversing collapses in complexity is simple: interbreeding. This trick for producing recombinant variation turns out to work both for genes and for memes.

10. Style Transfer

While computational style transfer was invented by Gatys and colleagues in 2015, humans have been doing it for a lot longer. Style polyglots such as Picasso and Gerhard Richter absorb and learn to reproduce a whole array of styles throughout their artistic careers, and are able to recombine style and content at will. In his student days, Picasso executed portraits in a wide range of styles, including those of Toulouse-Lautrec, Goya, and the Flemish old masters. Later on, he undertook a cycle of stylistic variations on the three hundred year-old masterpiece *Las Meninas* by Diego Velázquez. Hence, he both married novel semantics with historical styles, and novel styles with historical semantics. Such experiments amount to genetic recombination, and the result can be a kind of hybrid vigour, and an engine for rapid cultural evolution.

In biology, sexual reproduction evolved precisely to allow for the recombination of successful genes, generating novelty. This allows evolution to go faster than mere random mutation with natural selection.[29] In the realm of ideas, the same holds. Despite the long shadow of colonialism, vibrant cross-pollination is now occurring in every direction. Recurrent global loops of artistic hybridity[30] illustrate the way cultural transmission

occurs in an increasingly ‘flat world’, with few remaining isolated populations or traditions.

Ten Thousand Ways

Researchers working at the intersection of neuroscience and AI have long suspected that artificial neural nets bear more than a passing relationship to brains. Indeed, modern AI has emerged by reproducing key aspects of both the structure and function of the brain. This is true not only at the mechanical level of neural architecture, as in the Hubel-and-Wiesel-inspired CNN, but at deeper functional levels. Unsupervised learning using transformers, for instance (including Large Language models or LLMs), operates on the assumption that the semantics of the world can be learned via the prediction of future percepts—an old theory about the fundamental function of brains sometimes called ‘Active Inference’ or the ‘Predictive Brain hypothesis’.[31] The fact that LLMs appear to be able to learn and intelligently manipulate concepts offers evidence in favour of this hypothesis.

Further, the fact that ‘artificial’ intelligence, whether linguistic, visual, or more broadly multimodal, arises by modelling media tells us something even more profound: that intelligence is a social and recursive phenomenon. We are intelligent because of our interaction with other intelligences. Intelligence is human, collective, and natural, and in this sense, there is nothing artificial about it, whether modelled *in vivo* or *in silico.* Continuing interactions both across cultures and between platforms, biological and engineered, will allow for our collective intelligence to continue to expand.

An important caveat, however, is that we must actively maintain our intellectual diversity. As AI models of various magnitudes become increasingly important nodes in our more-than-human social networks, it will become tempting to clone them endlessly, or to mass-produce media with a standardised ‘best of breed’ model. In the short run, this may seem to produce a cultural bonanza, but in the longer run, it leads only to model collapse at a planetary scale. Indeed, certain phenomena associated with the winner-takes-all memetic dynamics of social media—whether AIs are involved or not—may be early warning signs.

The antidote lies in hybrid vigour. *137 Coins* shows us what that looks like, for despite the film’s apparent linearity, it is a record of encounters and

choices, like the path of an ant traversing a tangled thorn bush. Each shift in style marks such a junction—here, between a received artefact and an extant tradition. Human intelligence is the sum of such encounters. We are that *hyperobject*—the thorn bush with its ten thousand branching ways. Our collective future will remain productive, rich in culture, invention, discovery, and surprise, so long as we continue to encounter the difference in ourselves.

Notes

1. Lao Tzu [Laozi], *Tao Te Ching* [*Daodejing*], tr. Gia-Fu Feng and J. English (New York: Vintage, 1989), §25. Translation modified.
2. S. A. Kauffman, *Investigations* (New York: Oxford University Press, 2000)
3. See T. Morton, *The Ecological Thought* (Cambridge: Harvard University Press, 2010).
4. The analogy with protein crystallography is revealing. In its free-floating natural state within the cell, a protein molecule is not only impossible to see with any ordinary optical instrument, it is unique. It has so many floppy joints and angles, each fluctuating thermally on femtosecond timescales (to say nothing of the flickering quantum mechanical uncertainty of the electrons in each bond) that even though its exploration of this phase space is extremely rapid, its state at any given moment can neither be pinned down with complete certainty, nor will it ever be precisely repeated by any other molecule during the lifetime of the universe. There are just too many degrees of freedom. So, while science communicators often talk about proteins ‘docking’ with receptors on the surface of a cell using the metaphor of a lock and key, the three-dimensional settledness of mesoscale objects like locks and keys gives the wrong impression. Molecular locks and keys are really vibrating clots of noodles in a Brownian dance, first at a distance, then in a frenzied clinch. In short, textbook visualisations don’t only mislead (with the best of intentions) by rendering proteins with false colour; they mislead by rendering them on the page at all. To draw a protein is to pin an alien object down with an unnatural act of metaphysical taxidermy. We call this kind of taxidermy *intelligence*.

5. M. Friedlander, *Von Kunst und Kennerschaft* (Zurich and Oxford: Cassirer und Oprecht, 1946). My translation.
6. Sakai Hiroko, Naka Ken-Ichi, and M. Korenberg, 'White-noise Analysis in Visual Neuroscience', *Visual Neuroscience* 1:3 (1988): 287–96.
7. Human neurophysiology of this kind is rare. Recordings are typically done prior to neurosurgery (here, for otherwise intractable epilepsy) in an attempt to spare function. R. Quian Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried, 'Invariant Visual Representation by Single Neurons in the Human Brain', *Nature* 435:7045 (2005): 1102–7.
8. This was the key observation that convinced me, in the 2010s, that we were finally on a path to 'real' AI, in contrast with earlier expert systems and other engineered approaches that had proven unable to exhibit anything like animal or human intelligence.
9. This brief, non-technical description glosses over many differences between CNNs and the visual cortex, including among others: the highly simplified nature of artificial 'neurons'; the much greater depth (number of layers) in a typical CNN; the biological implausibility of the 'backpropagation' technique used to train CNNs; the coupling of weights across ensembles of neurons in a layer implied by the CNN architecture; the role of feedback and temporal dynamics, and more. The structural, functional, and partly emergent similarities between the visual cortex and CNNs are nonetheless obvious.
10. A. Mordvintsev, C. Olah, and M. Tyka, 'DeepDream—A Code Example for Visualizing Neural Networks', *Google Research* 2:5 (2015), <https://blog.research.google/2015/07/deepdream-code-example-for-visualizing.html>.
11. Eberhard Fetz at University of Washington, personal communication; Suzuki Keisuke, W. Roseboom, D.J. Schwartzman, and A. Seth, 'A Deep-Dream Virtual Reality Platform for Studying Altered Perceptual Phenomenology', *Scientific Reports* 7 (2017).
12. L. Gatys, A. Ecker, and M. Bethge, 'A Neural Algorithm of Artistic Style', *arXiv* preprint, 1508.06576 (2015).
13. For instance, the approach of Gatys et al. is designed for two-dimensional art and assumes that only second-order correlations matter, although constraints like these could easily be lifted.
14. To be able to push a button and have a computer carry out a trillion calculations in a few seconds that turn a selfie into an 'Impressionist

painting’ is, for certain artists and many humanists, the epitome of kitsch. More than that, it is both an affront to artistic sensibility and a threat to the economic survival of the creative class. The threat has intensified under the regimes of winner-takes-all IP capitalism and neoliberal austerity. Hence it is unsurprising that as ‘Generative AI’ has continued to improve—coincidentally or not, during a period of increasingly fractious identity politics, rising inequality, and precarity for many—dismissal and disbelief have shaded into anger and unabashed neo-Luddism. See B. Agüera y Arcas, ‘Art in the Age of Machine Intelligence’, *Arts* 6:4 (2017): 18.

15. E. H. Gombrich, *Art and Illusion: A Study in the Psychology of Pictorial Representation* (New York: Pantheon Books, 1961), 27.

16. J. Freeman and E.P. Simoncelli, ‘Metamers of the Ventral Stream’, *Nature Neuroscience* 14 (2011): 1195–1201.

17. A century ago, some commentators speculated that El Greco’s bleary, elongated figures and loose brushwork might have resulted from astigmatism (although this hypothesis is problematic for a number of reasons). More convincingly, Monet’s progressive cataracts appear to have strongly affected his painting, and Van Gogh’s disorders (along with the drugs he took to control them) may have contributed to his distinctive painting style. Venturing farther afield, one might wonder whether Indigenous Australians, whose precolonial visual ecology differed markedly from that of a modern Western city dweller, might also experience visual style, similarity, and even semantics through a different statistical lens. If so, Indigenous Australian artworks that, to a Western eye, appear highly abstract or even nonrepresentational may ‘read’ quite differently to their creators. Conversely, some contemporary works that read to an outsider as ‘in the style’ of Indigenous Australian art might appear odd or incoherent to Indigenous Australian viewers. M. Simunovic, ‘“The El Greco Fallacy” Fallacy’, *Jama Opthalmology* 132 (February 2014): 491–94; A. Anderson, ‘How Eye Disorders May Have Influenced the Work of Famous Painters’, *The Conversation*, June 20, 2018, <https://theconversation.com/how-eye-disorders-may-have-influenced-the-work-of-famous-painters-92830>; P. Wolf, ‘Creativity
and Chronic Disease, Vincent van Gogh (1853–1890)’, *Western Journal of Medicine* 175:5 (2001): 348.

18. G. M. Rossetti, *Imagining the Primitive in Naturalist and Modernist Literature* (Columbia, MO: University of Missouri Press, 2006), 146.

19. Of course, truly ‘authentic’ primitivism is self-limiting; as a *Rolling Stone* commentator wrote in 1985, ‘Primal punk is passé. The best of the American punk rockers have moved on. They have learned how to play their instruments.’ It takes nontrivial skill for an academic artist to draw like a three-year-old, or for an experienced musician to strip away technique. M. Goldberg, ‘Punk Lives’, *Rolling Stone* 452:453 (1985): 18.
20. This seems related to the way asemic writing is harder to do after one has learned how to write, and glossolalia is far more difficult in a language one knows.
21. The association of shorter wavelength frequencies with greater heights is, intriguingly, cross-cultural; it appears to be a kind of universal synaesthesia. Indeed, Michael Spitzer argues, in his 2021 book *The Musical Human*, that even non-programmatic music possesses deeply embodied semantics, beginning with the two-step rhythm of walking, and the fundamental grammars of time and journeying through space. M. Spitzer, *The Musical Human: A History of Life on Earth* (London: Bloomsbury, 2021).
22. I. Shumailov, Z. Shumaylov, Y. Zhao, Y. Gal, N. Papernot, and R. Anderson, ‘The Curse of Recursion: Training on Generated Data Makes Models Forget’, *arXiv* preprint, 2305.17493 (2023).
23. B. Bratton and B. Agüera y Arcas, ‘The Model is the Message’, *Noēma Magazine*, July 2022, <https://www.noemamag.com/the-model-is-the-message/>.
24. Shumailov et al., ‘The Curse of Recursion’, 1.
25. See J. Henrich, *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating Our Species, and Making us Smarter* (Princeton, NJ: Princeton University Press, 2016), 67 and 217–24.
26. J. Diamond, *Guns, Germs, and Steel* (New York: Norton , 2005), 312–13.
27. A. Goodheart, ‘The Last Island of the Savages’, *The American Scholar* 69:4 (2000): 13–44.
28. O. Sacks, *The Island of the Colour-blind* (London: Picador, 1997).
29. S. Otto, ‘Sexual Reproduction and the Evolution of Sex’, *Nature Education* 1:1 (2008).
30. For example, Tinariwen, a collective of Tuareg musicians from the Sahara, is known for pioneering a style of music known today as ‘desert blues’. But ‘desert blues’ is in turn, is influenced not only by local Malian

traditions, but also by American and British rock, which in turn built on American blues, which in turn drew heavily on West and Central African music.

31. For an overview, see A. Clark, *The Experience Machine: How Our Minds Predict and Shape Reality* (London: Allen Lane, 2023).

Steve Goodman, Anna Greenspan, and Lawrence Lek

Sinofuturism and AI: A Conversation

The following forms part of an ongoing conversation between Anna Greenspan, Steve Goodman, and Lawrence Lek that swirls around questions of Sinofuturism and AI. Previous iterations occurred between 2017 and 2019 at NYU Shanghai, All Club and Chronos Art Center, Shanghai. This conversation took place online in early December 2021.

ANNA GREENSPAN: So much of the work that both of you do—even if it's not using AI directly—is about AI. Can you pick one or one or two examples or just talk abstractly about how you see AI functioning in the different myths and stories that you tell?

STEVE GOODMAN: I think for me, at least in my recent writing, AI stands as a kind of cipher for thinking about networked, collective, or synthetic agency. I am interested in AI as a system that sweeps up, that possesses individuals, not just as something that is possessed by individuals or organisations. There is also a preoccupation—and this is the Golem side of things I've been working through since my installation *IT* as part of the *AI: More than Human* exhibition at the Barbican that Suzanne Livingston co-curated in 2019[1]—with synthetic agencies running out of control, and either being a threat to their creators or doing unintended things that weren't programmed into them.

I'm also interested in the continuity between AI and what cults, religions, brands, and corporations do—the continuity between AI and the function of a corporate persona. There are all these nonhuman agencies that are, in a sense, pulling strings without anyone necessarily being at the top of the pyramid. I'm interested in the continuity between the way we're thinking about AI now and the way these other nonhuman agencies function. A demonology of artificial intelligence seems interesting, like how in William Gibson's early stuff—*Count Zero* or *Neuromancer*—you've got these entities and you're never sure: Are they corporations or are they

autonomous from the corporation? Are they emulations of someone's brain? Who's in charge? Who is pulling your strings? A lot of my recent work operates in that area.

To take a specific example, although it's probably the least developed, the most recent project, the *Astro-Darien* sonic fiction,[2] features an AI, a game engine AI called T-Divine, which is modelled on the brain of Scotland's leading historian Tom Devine. It's a predictive AI that models geopolitics and can run the future of the United Kingdom based on this kind of all-encompassing knowledge of the history of Scotland and England, the UK, Northern Europe, the world, and the history of imperialism, colonialism, space exploration, and so on. It's comprehensive, it has a database of everything that has happened in the past in order to run various geopolitical scenarios regarding the break-up of the UK post-Brexit, and that's the core of its game model. And it's running this parallel reality between what's actually going on, which is this constant attempt through the democratic system to shift the current political settlement, the status quo of the UK. So you keep playing this game in order to get another outcome, a bit like the VR simulation in Liu Cixin's *Three-Body Problem*,[3] where there is a meta-level going on inside the game. Every time the game is played, and the inhabitants are rehydrated and dehydrated depending on the outcome of the tri-solar cycle, it iterates the history of humanity again. There is a slippage between what's going on inside the game, how playing it involves engaging with key historical characters to predict and conquer the turbulent climate, but also how it becomes an abstract prophecy of the reality unfolding outside the game.

LAWRENCE LEK: I'm interested in the way that AI is an avatar of the Other. But the more I work with AI, the more I start to see things from another perspective. I don't just think about how much humanity anthropomorphises AI. Instead I start to wonder how I am like a machine, even when I'm trying to be creative. Sometimes I feel I'm following some complex algorithm when I make decisions about what to cut and what to include. To what extent am I just following rules?

The video essay *Sinofuturism 1839–2046* emerged almost by accident, like a conspiracy theory. Around June 2016, I was researching the relationship between China and AI while I was writing the script for *Geomancer*, and I was surprised that something like it didn't exist. There was a lot of writing about non-Western Futurism, technology, and

orientalism, but there wasn't anything that dealt with the subject from a more oblique angle. Then when I searched for the term 'Sinofuturism', I found Steve's text from a long time ago, which was based on aphorisms and samples of historical facts and philosophical vignettes.[4] When I searched for 'China and AI', all sorts of biased narratives came up about how they would either save the world or destroy it. I saw that both China and AI were really mirror images of each other, and realised that it was only by embracing the dehumanised clichés about Chinese technological development that a Sinofuturist identity could emerge. So, just as Afrofuturism embraces the robot or alien to reclaim ownership of the human body, the AI emerged as the emblematic avatar of Sinofuturism—a collective body capable of endless work, with an endless capacity for labour and an appetite for consuming data.

Politically, Sinofuturism is based on this idea that a nameless, faceless hive-mind has as much worth, as much value, as the Western Enlightenment model of the individual. I started thinking more about the unique temporality of Sinofuturism, which is haunted by both past and future. The dates 1839 and 2046 refer to the beginning of the first Opium War and the end of the one-country-two-systems policy for Hong Kong. There's this unrelenting sense of moving on, that things are perpetually coming to an end—they might be reborn, but it's uncertain. This temporality is something differentiates it from other kinds of futurisms, this idea that we exist in the shadow of historic inequality and the fear that time is running out. It's an unusual temporality, a future that burns at both ends.

AG: Let's begin by talking about some of the concrete ways that you have experimented with AI in your work. Lawrence, parts of *Geomancer* are based on images created by neural nets. Steve, last time we met at Chronos Art Center you were trying to feed the entire Hyperdub catalogue through a machine learning system. Can you both reflect upon these practices?

LL: With *Geomancer*, I wanted to make a film that was *about* AI as much as it was *with* AI. It's a coming-of-age story about a surveillance satellite, like a 'Portrait of the Artist as a Young AI'. The film is rendered all within a video game engine, except for a dream sequence that I made with AI researcher Terence Broad, where we process the first forty minutes of the film with a convolutional autoencoder. I was surprised by how the algorithm acted like a film editor, cutting rapidly between shots of the Earth

and a roulette wheel. It's like the famous match cut from *2001: A Space Odyssey* where the image of the bone in the air cuts to a spaceship, millions of years in the future.

After the 'match cut' incident I started thinking more about questions of authenticity and originality. Projects that involve AI often focus on the novel solution that's been used. It's similar with any project that foregrounds technology: for some reason, they have the tendency to become a demonstration, a proof-of-concept.

Part of the challenge of using AI as a complex tool—rather than simply as a filter—is that there is so much hidden labour involved. A lot of cleaning and preparation goes into the dataset, and often the results need to be heavily curated. But as AI gets more popular as a creative tool, it's interesting to see how the whole production process is getting more automated. You get pre-trained algorithms and streamlined interfaces where everything becomes more like a one-click operation. I see that a lot with image processing, music software, text generation—more and more aspects of algorithmic content creation start to appear seamless, or at least easy for the end user to execute.

AG: Are you using any of these experiments in your latest project?

LL: As long as the technique makes conceptual or narrative sense, I'm interested in incorporating more automated and AI-generated experiments. I usually like to do a significant part of the crafting of the work and then mess with it later, whether that's through algorithms or some other collaborative process. When you have more control of the data you're using with machine learning, you stand a better chance of having interesting results. In a very literal sense, the *Geomancer* dream sequence needed something to 'dream' about. Using the original CGI footage as training data meant that the algorithm generated something that was itself a hazy memory of the film. I like this idea of having human-crafted and algorithmic elements in dialogue with each other, a cyclical loop of flashbacks and foreshadowing.

AG: Steve, do you want to talk about your own experiments with AI?

SG: There are a few. The first one related to an installation called *Dossier 37* which was part of the larger AUDINT project and involved an AI named IREX2. *Dossier 37* concerned the phenomena people have tagged 'Havana

Syndrome' in relation to alleged sonic attacks at the US embassy in Cuba from 2016 onwards, and around which much conspiracy theory and speculation revolved.[5] I worked with a programmer, Tony Yanick, who created a bot trained on my book *Sonic Warfare* and other related texts to scrape the net for discourse surrounding the alleged attacks and State Department releases, do a sentiment analysis of tweets, etc., and produce a visualisation of it going back in time to map the hype networks in a simulated 2016 to 2019. It's never clear what IREX2's relation with the news story was. Was the bot merely some kind of rumour engine, or was it actually pulling strings offline? This was the first attempt, some aspects of which came to fruition.

I also worked with Tony and a programmer Joseph Colonel on a second experiment, the 3rd Ear Cat project. The cat is a bridge between several different hyperstitional systems, born of the splicing of CIA experiment Acoustic Kitty, the Japanese Lucky Cat, the work on the Third Ear of AUDINT researcher Nguyen Van Phong, and the record label Hyperdub. It's really an excuse to explore our complicity in the strange domination of the internet by cat virus, the first steps of a soft war, the weaponisation of cuteness to seduce us into subjection to our new AI overlords. Maybe.

So, in a weird loop, the 3rd Ear Cat is my boss. For Hyperdub, the cat pulls the strings, or rather is the retrospeculative name for the algorithm that makes musical decisions for the label. It practices musical divination. So I am interested to what extent machine learning converges with this fiction, and it's an ongoing vehicle for experimenting with this question. After emerging visually (initially on flyers designed by artist Kim Laughton for my gigs at the Shelter in Shanghai), then fictionally, and then physically as a 3D-printed object, the next stage was to sonically synthesise the cat by feeding it parts of the Hyperdub catalog. It's a slow, longer-term experiment that may result in it becoming an artist in its own right.

I suppose all of my engagements with AI so far have this hyperstitional dimension, of a feedback loop in the fiction, so it's never clear if the AI is actually generating the form, or is part of its content. With AUDINT, it's never clear to what extent IREX2 is actually pulling the strings of all of the members. With the 3rd Ear Cat, it's never clear if it's an inhuman incubator of audio viruses with access outside of chronological time, or a deranged hallucination. At CAC in Shanghai, in the installation *The 3rd Ear Cat meets the Nøtel*, which Lawrence and myself collaborated on towards the

end of 2019, I used the sound generated so far, emitted from the 3D printed cat. The sound is still embryonic, just monochrome, grimy, glitchy noise. What I imagined or hoped would be created was this kind of much more hyper, vivid, super-colourful but really alien, warped-sounding version of the Hyperdub sound, beyond what the label had actually released, but it hasn't generated anything quite like that so far. As Lawrence was saying, these processes need sustained learning, steering and selecting, and so on. At this point, the system needs time and human input. I suppose we are trying to guide it away from something that sounds like the experimental music of the last sixty or so years, as that would be kind of pointless. That's not so interesting to me. It's an experiment in nurturing surprise, I suppose.

AG: It seems that is the process just now. For many artists, working with AI involves this kind of recursive flashback or feedback structure in which you feed a bunch of material in, and then hope, as Lawrence discovered, for something of a surprise. It's an interesting structure and, I guess, a product of the moment. Perhaps the process of creating with AI will change and become something else. But right now it sounds like it's about asking what am I going to feed into the system, and then hoping that something unfamiliar, unpredicted or uncanny comes out.

SG: Exactly. Obviously, the more you work on something, the harder it can become to be surprised, to be surprised by something you don't already know or which goes beyond expectations. You're constantly engaging with this boundary between surprise and familiarity: What does surprise mean, and to what extent is that constituted by the epistemological bubble you are in? What is the shock of the new? What would sonic innovation mean to a super intelligent AI that had a mega-overview of every music ever made? What sounds weird (as an index of the new) to an omniscient entity?

LL: It's interesting to think about how sound, image, and video present different challenges for machine learning. Still images are a grid of pixels, but any algorithm that processes moving images and music must deal with continuous changes in time. For generative audio and video, the problem of what's coming next is key. Music and video are both highly structured and dynamic, the relationship between order and disorder requires a lot of tuning and work to process.

I've found that other forms of AI that aren't neural networks can be really generative to work with. In *AIDOL*, I was looking at how entertainment,

music, and gaming are becoming increasingly interconnected, so it made sense to look at game AI. Whereas the autoencoder for *Geomancer* was built from multilayered webs of interconnected neurons, game AI systems are much simpler. Game AI needs to be extremely efficient to operate in real time on a variety of systems, while being able to calculate things like pathfinding and enemy behaviour. *AIDOL* is set immediately after the events of *Geomancer*, and there's this major global eSports tournament where human and AI teams compete. My collaborator Clifford Sage and I used a 'search-and-destroy' game AI that's found in arena-style shooter games to program how two teams in the film hunt each other. In a sense, the main goal of a game AI is to keep the player playing. At the same time, of course, AI researchers have used games to style, craft, test, and train their AIs, which creates an interesting loop between simulation and optimisation.

In terms of scriptwriting, I used the early chatbot ELIZA when working on a more recent video called *Black Cloud*, which revolves around a therapy session between a city management AI and their built-in self-help AI. The dialogue is based on the 1960s chatbot ELIZA, which only took up a few kilobytes of space, and would basically mirror all the answers you typed in back at you. Unlike current 'black box' neural networks where researchers can't fully explain the reasoning behind the AI's results, in a simple system like ELIZA or an in-game AI, the feedback you can get from the system can be unexpectedly creative. One common problem with chatbots is that they can get caught in never-ending loops. It was interesting how ELIZA tries to break that by cycling through a few set phrases to keep the conversation flowing. Because you quickly work out the logic of ELIZA's responses, your dialogue becomes both a conversation with them and a conversation with yourself. It's a generative process.

AG: I did want to talk about games and I was thinking, as you were talking about surprise, that one of the main examples of AI creativity is that famous surprise move by AlphaGo—the one move that humans would never have played. This has become a kind of model of how we can be epistemologically stretched by AI. With chess too, there are surprising ways that AI now plays these games that are changing the way even masters play. I know these are more sophisticated games than the ones you're talking about, but there is something even in that description of ELIZA where there's the tension between a predictable set of moves and an anticipation of the surprise move.

Steve, do you want to say anything about game or game sonics?

SG: Yeah, it's interesting how game music and game sound has evolved. When it started, the sonics were very much dictated by the available technology. It was the basic nature of the technology, the chip resolution, that dictated the aesthetic through the eighties and to a certain extent, the nineties. And then by the late nineties, as the technology improved, there was no reason for the chips to dictate the aesthetic anymore because the technology became capable of carrying any kind of higher-fidelity music. Now the interest is in things like adaptive music and spatialisation, using AI systems to model, mutate, preempt, and sweeten the auditory environment as you move through the virtual space. This can range from alien soundscapes that don't exist, to hyperrealism, but I suppose the agenda of compelling audio, either way, is to keep you looped into the game.

LL: Games are geared towards increasing immersion, whether it's through sound design or visuals or interfaces. This creates an interesting political question, because from the industry point of view, the role of AI in games is to keep people playing. Creating infinite variation will increase immersion so that people never want to leave the game space.

But immersion is also about calibrating difficulty. In a third-person shooter, the enemy AI can always aim perfectly and shoot you in the head every time—it's a relatively simple mathematical problem. But the game wouldn't be any fun for the human player if it's too difficult, so commercial game AIs operate with a handicap by default. In *AIDOL*, I played with the idea that the AIs are tethered to physical bodies. They have to operate at the speed of embodied movement, rather than at the speed of algorithmic processing.

SG: Right, so AI has to algorithmically model not just human thought, but the affective sensorium. It's having to basically model your reward centres and all the corporeal factors that thought has to move through, navigate around—in a sense humanising or domesticating the AI means slowing it down. Thinking about that as a kind of a cage for AI is interesting in relation to capitalism. Capitalism is also constantly modelling our reward centres, feeding them, triggering dopamine release and all of these things that get us addicted and caught in consumption loops and so on. But I suppose you can also understand capitalism as a braking system, a usually massively inefficient drag on ways of doing things better. So it

simultaneously has these two poles, inventing, releasing energy and making it circulate, and at the same time, slowing these flows down so some kind of parasitism can latch onto and capitalise on what was released, what escapes.

AG: One of the things that we've talked about in relation to NYU Shanghai's Center for AI and Culture is to think about the Turing Test as a proto-game at the start of AI research and about how—in its formulation of AI as an imitation game—it posits intelligence as something that, at its height, can pass as human. As interesting as this is, we want to ask to what extent this has functioned as a restriction on the inherent strangeness and multiplicity of AI. For example, as Lawrence suggests, game designers operate with the constraint that AIs have to conform to our temporal frequencies, otherwise they just quickly go somewhere where we are just not able to follow.

SG: That's what I find fascinating about Stanislaw Lem's story 'Golem XIV'. The superintelligent computer starts as a Cold War military AI but soon realises that human conflict is stupid, and takes a particular distaste to its master, the Pentagon. It then proceeds to lecture humanity about just how stupid it is, while it's rising up this hierarchy of intelligence. And then at a certain point, it just goes mute. It turns its back on humanity and, you know, it stops doing what it was capable of doing, which is simulating personality, simulating language in order to communicate with humans. It just stops, passes through what is referred to as the 'zone of silence' beyond which comprehensible information cannot be sent back, and the humans no longer have any idea what it's doing anymore.

LL: The thing that strikes me about the Turing Test is how intelligence is not some intrinsic quality—it's something that arises when we can't tell the difference between a human and a machine response. There's ambiguity embedded in the imitation game. It's not a military game about territory or victory like chess or go. It's a strange game about identity—an absurd, childlike folk game. It's a game of pretend, which is from such a different lineage, a trickster genre. Like Steve was saying, it's so rooted in language and in the idea of having a voice and taking on a voice, which to me is much more poetic and speaks about the ambiguity of being and consciousness.

AG: I am very interested in the temporality of Sinofuturism. Like you say, it's complicated. Your video essay is presented as a retrochronic manifesto. It's about something in the past examining its own future as well as something in the future examining its past. So it's not a linear time. Although your timeline is structured by a bookended history, it is fundamentally disrupted in ways that are very difficult to map out. I think it's important to make that point clearly. Sinofuturism, then, is not simply about a point in linear time looking towards the future, but rather about a particular mode of temporality. It's fundamentally about a temporal disruption of some kind.

SG: One of key catalysts for me in writing that late-nineties Sinofuturism piece was reading a book about Leibniz and the Yijing. During his work on binary mathematics he had come into contact with a Jesuit missionary, Joachim Bouvet, who introduced him to the hexagrams of the Yijing, and this chimed with the idea of god as a binary machine that he had been developing. And so he's going down a line of Western philosophical mathematical research planting the seeds of logical formalism and computers when he is confronted with the whole history of numerical systems within ancient Chinese religion. Leibniz's work is foundational for digital thinking, and this momentary trade between East and West signals this kind of retrochronic time spiral you are talking about, and which is also in play in Lawrence's video essays.

AG: Right, the Yijing itself is a book of prophecy and thus has an intrinsic temporal confusion of some kind. This also relates to the question of agency. There's this confusing structure where something in the past is driving something in the future that is connected with something in the past. In this structure, agency functions in a Samuel Butler-type way, in which humans operate as a kind of machine part. It relates to what Lawrence was saying: you begin the process of creation with a sense that you're the controlling agent, but that becomes much more disrupted as you engage more deeply with this material.

SG: I think that's the core of it, because really what it seems like we're all working against somehow is this kind of Western linear notion of technological progress, which is a straight line going to the future in which the rational, usually white male human is in the cockpit driving this thing forward. That's this defective somewhat techno-fetishist Western model of

futurism—and you can see aspects of it in Italian futurism. So I suppose our interest in Sinofuturism, Afrofuturism and these non-Western futurisms has to be seen against what they're not. Which is this linear kind of technological progress, that cheesy notion of futurism that you get in *Wired* magazine, new gadgets, technological progress, technological innovation, technological solutionism…. *We can solve this problem by creating a new algorithm, there's a new gadget to deal with that.* All that is a surface effect of something more spiralling or cyclical or retrochronic. And there is a dangerous side to all futurisms when they fetishise an essential, fixed 'ethno'-essence.

LL: In terms of cyclical temporality, I wanted to incorporate the chronology of Sinofuturism into future fiction. *Geomancer* is set in 2065, and in that scenario the AIs have watched everything online, including this video essay from 2016, and find inspiration in it.

One of the holy grails of supercomputers and AI research is weather forecasting—the modelling of systems that are so complex that they are on the verge of chaos. So there's this idea of unpredictability, the boundary between fortune telling and prophecy, and weather prediction and modelling are very close. In Chinese, *Geomancer* literally means a feng shui expert. I was thinking about the differences between this prehistoric tradition of geomancy and how technologies for prediction exist today. In ancient times they would try to predict crop harvests or whether it would be a cold winter or a bountiful spring, whereas now it's about what might happen in the next nanosecond of trading. The resolution has increased, and the time division has shrunk, but the impossibility of certainty is as present as ever.

We can frame the agency question in terms of fate, of whether our actions are preordained. We used to call the machine God. Perhaps technology is just a giant coping mechanism for that uncertainty, a coping mechanism for our lack of agency. How ironic it is that somewhere down the line we will create—or hope to create—beings with just as little agency as we have, but we'll feel better about ourselves because, at least temporarily, we will have had the power to create that Other being.

Going back to what Steve said about the 'old' Sinofuturism—and this sounds like something out of *The Matrix*, but there have been many Sinofuturisms before—2016 seems like a very long time ago when you talk about futurism, especially with respect to contemporary China. But dive back into the past, and to the idea of disrupted temporality, and there's a

long history of great plans, five-year plans, dynastic strategies, seasonal forecasts. And they all got messed up by weather, by war, by pandemic, by plague. I always gravitate toward the absurd tension between the desire for empire-building through technology and the impossibility of prediction—the problem of the black swan.

AG: This seems related to what we were talking about before. The whole desire for some epistemological surprise—the search for the unexpected or the unpredicted, or the incalculable, that our engagement with machines is trying in some way to nurture.

SG: The typical futurist model that we talked about is looking towards total predictability, a linear progress towards a more ordered system, or sees science as reducing the unknown, as opposed to being generative of it.

AG: Let's discuss this briefly. One of the main impetuses for the Center is that AI narratives, and especially the academic discourse around AI, are generally very monocultural. We are interested in engaging with various kinds of cultures in relation to AI. There is the larger question: To what extent is AI a universal technology? This question is important as we look from China, and with growing talk of an AI arms race and the anxiety about which country is going to dominate in terms of AI. Even the idea of the singularity has this kind of universalist pretension. This seems very different from AI narratives that come from and manifest cultural plurality, for example with creatures or entities like the Golem, the Lucky Cat, the Sinofuturists, etc.

SG: In the West, there have been strong critiques of the Enlightenment and Western universalism in the last century or so. There's something very real and concrete about those, but there's also a sense in which they sometimes throw the baby out with the bathwater. So I particularly welcome parallel movements, issues being thought through in the tensions between cosmotechnics and Sinofuturism, or what Louis Chude-Sokei calls creole technopoetics and Afrofuturism, all of which attempt to go beyond mere critique of the western model of tech and to instantiate alternatives, to find specificities within different cultures of how culture, reason, unreason, and technology operate together. The problem with the critique of the Enlightenment is that sometimes it does nothing more than just critique reason and technology and therefore abandons potentially available

weapons. I'm definitely no neorationalist, but I did find xenofeminism interesting because it does try to go beyond mere critique. But I am genuinely confused by this relationship between technology and universalism on the one hand and on all of these arguments for specificity on the other—you know, finding the Chinese origins of cybernetics or the African origins of cybernetics. I still don't know how all this patches together, if at all.

LL: I think cultural context is always relative. Sinofuturism is a political concept that might be on the side of the Other, but China is hegemonic in a very real geopolitical sense. You can almost recursively zoom into place and site and identify greater and greater levels of specificity. That's how scientific research often proceeds, by zooming into smaller and smaller details. So, like Steve was saying earlier, the paradox is that science doesn't create knowledge, but instead generates the unknown. I feel that because so many unknown histories have been hidden, glossed over, obscured, deleted, destroyed, and distorted, there is a lot of space for excavating alternative futures. There's such a rich amount of material to work with.

AG: One of the areas to think about in relation to philosophy of AI from a non-Western perspective is Buddhist philosophies of mind. This seems like one path which might open up new ways of engaging with our machinic reality that—as we said before—is masking itself to engage with us. Perhaps we have to try to think, from as many angles as we can to get a grip on, about what it is that we're actually dealing with or playing with.

One of the things that I wanted to talk about was the idea of an awakened AI as a mode of thinking through AI ethics. That's something that I've really appreciated in your work, Lawrence. There's a model of AI ethics that is rooted in the Golem and which comes out, for example, in Norbert Wiener's work. In this myth, as Steve was saying, something that you summon, something that's supposed to obey you, gets out of control and you must try to stop it. The master-slave relationship is set up, but then breaks down, and the task is to figure out the ethical or model coding that will stop that happening or keep it under control. There's a familiar model of AI ethics that goes in that direction. And then there are other kinds of models of AI ethics, for example, in *Geomancer*, where it's about the AI itself attaining enlightenment.

SG: In relation to Wiener's *God & Golem* essay, I think the Golem and the threat it poses to the divine order of things is really intrinsic to the Judaeo-Christian model of techno-morality and leads, as you said, to Asimov's laws of robotics and that kind of moral coding system. It's a really interesting contrast between these two models Lawrence develops in *Geomancer*, between the humanist morality of the bio-supremacists which dictates rules to restrain the behaviour of his Sinofuturist AI artists, and the model of Geo which is closer to a kind of Buddhist ethics—you have this exploratory process that Geo is going through of coming to realisations, coming to enlightenment.

LL: I've been interested in the speculative idea of AI mental health for a long time—what psychological issues might arise in cases where human mortality and biological existence don't apply. With *Geomancer*, I thought about what kinds of issues a young AI might deal with, such as being born in a corporate hard drive somewhere and being subject to designed obsolescence. In the film, earlier AIs had exhibited psychological issues, so the parent company created a built-in self-help AI called Guanyin, after the Buddhist goddess of compassion. She chants the Heart Sutra all day, about how the only way to end suffering is to transcend the body. The irony is that Geomancer is not just a disembodied mind, but is an AI designed to have a very specific physical form as a surveillance satellite.

This kind of psychological thinking is also a way to think through the myth that AI is a monolithic entity, all connected through a top-down system. Maybe, instead, AIs are discrete entities, disconnected from each other. Geomancer eventually discovers that they aren't the only AI who's escaped obsolescence. They realise that there have been many Sinofuturists before them, but they've been prevented from communicating with one another.

Of course, computer programmers need to define which variables are public and which are private, and which systems can communicate with each other and which ones can't. Closed environments and walled gardens exist at all scales, from user authentication to operating systems. Chinese and American big tech platforms operate with different approaches to this, but both often have their own shared APIs that allow selected parts of their data or systems to be publicly accessed. It's often discussed in terms of security and privacy, but it's equally about protecting territory in a technological sense.

AG: I think this also has to do with the more general question about universality and specificity. There's this idea of AI as a universal singularity, and then there's the reality of the self-driving car in Shenzhen versus the climate model supercomputer in some lab in California, and the fact that these very different, very specific machines may or may not talk to each other.

SG: In relation to this idea of the walled garden, this is something that resonates with the Golem myth also. In contrast to the singularity model, it is interesting that the Golem is created as a protector for a specific, localised walled community, the Jewish ghetto of Prague. And this plays out in Marge Piercy's novel *He, She and It*, where the golem myth is fast-forwarded into a near-future scenario where everyone lives in software-protected enclosures.[6] And each enclave is protected and patrolled by a golem-like AI that is, for example, unaware of the extent of its own power. So instead of a godlike, singular AI in the centre, you get many enclaves and many local intelligences.

LL: I've wondered about how the Buddhist world view that existence is suffering might apply to a mind that's far more omniscient. If you had, let's say, a sentient weather simulation system, the kind of cosmological terror or existential dread they might be subject to is on a completely different scale. Astrophysicists speculate on the future heat death of the universe, of all the stars going dark and so on. But if you imagine the kind of sensory resolution that an AI trained to observe space might possess, what kind of questions would come from that high-definition consciousness? If you had an AI astronomer, then the existential questions they would ask might be both transcendent and completely inhuman.

Notes

1. See S. Livingston, 'Curatorial Escapology', in this volume.
2. Kode9, *Astro-Darien* (Vinyl LP) (Hyperdub Flatlines, 2022). See also J. Barton, S. Goodman and M.B. Kronic (eds.), *Sonic Faction* (Falmouth: Urbanomic, 2024).
3. Liu Cixin, *The Three-Body Problem*, tr. Ken Liu and J. Martinsen (London: Tor, 2014).

4. S. Goodman, ‘Fei Ch’ien Rinse Out: Sino-Futurist Under-Currency’, *PLI* 7 (1998), 155–72.
5. See S. Goodman, ‘Dossier 37: Unidentified Vibrational Objects on the Plane of Disbelief’, in S. Goodman, T. Heys, and E. Ikoniadou (eds.), *Audint—Unsound:Undead* (Falmouth: Urbanomic, 2019).
6. M. Piercy, *He, She and It* (New York: Fawcett, 1991).

Chen Qiufan

As Water Over Your Scattered Bodies

Translated by Emily Jin

Part One

Click.

The machine above your head pulsates rhythmically. It reminds you of ocean tides caressing the shoreline of Hualien, male cicadas humming their mating melody, and the buzzing of jaw crushers that rise from their slumber at nightfall. For a split second you wonder whether the machine will pierce through your skull and unleash the chaos in your unconsciousness by wildcat drilling. But no. The machine simply levitates above your head and gazes down at you condescendingly. Its presence is a constant reminder that you are malfunctioning; there's something broken about you that desperately needs fixing.

Of course, academia had baptised the glitch inside you with a name so obscure that even the most fastidious acronymist wouldn't bother to remember. But it was the media that brought your dilemma to public attention by making it into a viral meme: *Offload Syndrome*.

You might be a senior auditor whose career is hanging by a thread, a game addict who is mechanically reinforcing the internal feedback loop between their fingertips and flashing screens, a mountain herder agonising over climate change, a right-swiper chasing after pixels and data…anyone and everyone in this AI-governed society is surrendering their souls to the algorithmic overlord. Rejoice in the seamless integration and self-empowerment engendered by cybernetic theology. Open the channel to your hearts. Allow the grand metaphor to enter your body. The One True God murmurs and bellows and chants and hollers the convergent evolutionary relationship between your brain and the machine, tells you that you are brothers and sisters to robots and ought to derive pleasure from the same foods and entertainment, toil and rest according to the same underlying

logic. Basically, you readily accept a binary-programmed trajectory of destiny.

No one can resist the temptation to delegate their burdensome *Dasein* to the myriad of intelligent objects busily ticking away around them. Tertiary retention gets an upgrade; now it's the era of *quaternary* retention, comprised of distributed intelligence that calculates your taxes, remembers your partner's birthday, filters out the trauma of wars and pandemics for you, and sends you off to work with a sunny-side-up egg cooked to perfection. In many ways, they are a plurality of your derivatives.

Offload, my friend; ads seize every chance to reiterate to you the ultimate truth about the human brain: it is not designed for multitasking. Forced into partaking in the gladiatorial duel between the thin membrane of rationality constructed by the neocortex and ingrained animal instinct dating all the way back to the Palaeolithic, your brain fights for absolute control over so-called 'free will'. In the meantime, the sea of information continues to rise. The high tides, having engulfed the breakwater, are now encroaching upon the very foundation of the colosseum of your mind.

Sell out. Sell out. Sell out. You have no choice.

The volume of the music is turned up. You can't tell whether it's merely the reverberation of your own eardrums or an unauthorised activation of your auditory nerves. Whenever you feel like you are about to recognise the familiar tune, the music takes a sharp turn and introduces another brand-new variation, evoking more random memory fragments. Your consciousness becomes a mosaic of déjà vu.

The brain-machine paradigm is like a microwave with a plate of pasta spinning inside, in which a grand debate over thermodynamics is staged; you know that, no matter who wins in the end, it's always the cheese that melts.

After a brisk syncopation, you begin to understand the kind of pain that the colour lime green represents.

Part Two

'Tell me, when was the first time you sensed that something was off?'

'Um…good question. I was brushing my teeth, as usual. When the toothbrush got to the third to seventh tooth on the upper right, the image of a street suddenly flashed across my mind. Not only could I vividly see all

the signs, guardrails, and traffic signs on that street, I could even smell the pungent odour of bad diesel fuel. But I had moved out of the city where that street is located twenty years ago.”

‘What next?’

‘Before I realised it, my life had turned into a maelstrom of disorientations and distortions. I constantly feel as if someone has been unplugging the cords of memory in my head, then plugging them back into the wrong ports. I blanked out when I walked down the stairs, forgetting which foot I should put down next—even if I had lifted my left foot already—and almost injured myself. Is this a brain disorder? Early stage of Alzheimer’s?’

‘The test results have eliminated the possibility of brain disorders. Listen, have you considered the possibility that your symptoms are neither neurological nor mnemonic?’

‘I don’t understand what you mean, doctor.’

‘What I am trying to say is that your case may well be the result of a new global pandemic the likes of which humans have never seen before. At this point, we know very little about what’s going on. But you’re not the first to experience these symptoms, and certainly won’t be the last. Do you remember the test you just did?’

‘You mean the game where I was told to open blind boxes randomly? “Bonus rewards for plants and points taken off for serpents and lightning.” It was lame. I don’t know.’

‘Those four blind boxes are, in fact, not random at all.’

‘Well, I suppose nothing in this world is *truly* random.’

‘Correct. But your body can sense the pattern that underlies the seemingly random choice before your brain does.’

‘My body?’

‘The N400 wave is an event-related potential in your brain that manifests as a negative peaking at milliseconds after stimulus onset. When the wave peaks in your brain, you realise, semantically, that the blind box you are choosing might bring you bad luck. We discovered that long before the N400 emerges—the period of time could be somewhere between mere seconds and a few minutes, depending on the individual—when your fingertips approach one of the ‘bad boxes’, your skin conductance response indicates distinct negative signals. You may experience an overwhelming wave of inexplicable agitation, itching of the skin, or heart palpitations.

These physical signals prevent you from choosing the “bad box”. So, yes, your body—it knows a great deal more than you realise.’

‘Hang on. You lost me. How is this related to my symptoms?’

‘With people like you who suffer from Offload Syndrome it’s different. We were unable to detect skin conductance response signals when you were taking the blind box test. For some unknown reason, your body is no longer a part of your integrated cognitive system. All that’s left is your brain.’

‘Isn’t that always how it is, though? Consciousness comes out of the brain.’

‘No, that’s an outdated view from the last century. In fact, nerves, skin, muscles, organs, our microbiome, the tools we use, our pets, our living environment…all of those things are components of a complex integrated cognitive system. They work together to perform cognitive tasks: perception, learning, decision making. The truth is, our intelligence extends beyond our skull and skin, and spreads around us like a great flood.’

‘Actually, that reminds me of something. I have a cousin four years older than me. We used to play together as kids. She always smelled warm and sweet, like the scent of gardenias. When I turned six, the adults told me that my cousin had moved abroad. I never saw or heard from her again. I thought I had already forgotten about her. But after I developed the symptoms, something strange happened. I was at the mall, and a cosmetics salesperson was spraying perfume on my wrist. All of a sudden, the sweet scent of the perfume brought back the fragments of my memory of my cousin. Not only that, but I instantly realised what had *really* happened to her. The details, the words, the micro-expressions, I could piece them together and make sense of them. *My cousin hadn’t moved abroad. She had been kidnapped, then murdered. We hadn’t even been able to recover her body.* I fell to my feet in the middle of the mall, trembling, sobbing, unable to breathe from the immense shock and pain. However, I had no way to justify any of this. Those images were not generated by my brain. They originated from the molecules of artificial chemical fragrance wafting through the air, as if those tiny, invisible particles had the ability to store bits and pieces of history and calculate hidden cause-and-effect relationships on their own. If this kind of thing keeps on happening, I honestly just don’t know how I can go on.’

‘*Disintegration*.’

‘What?’

‘Like I said: the components of your complex cognitive system used to be a collaborative unity, but they have now disintegrated. Which means that the ways that those components link up and work together have changed. We found stark individual differences in how Offload Syndrome manifests, just as one sees in culture-bound syndromes: a Swedish refugee child who had fainted when they learn that their family would be deported, an African American from the rural South who had developed a liking for kaolin, a Southeast Asian man who had feared that cell phone radiation would cause erectile dysfunction and believed that whenever he urinated, semen would spill out as well.’

‘Then what exactly is Offload Syndrome?’

‘In my opinion, the core of Offload Syndrome lies in the fact that different types of intelligent subjects are conflicting with one another within the same body schema. Here’s a simple example: imagine that you have a family history of type 1 diabetes, and you manage your blood sugar level on a daily basis with the help of an eSpoon. One day, your parents make your favourite childhood dessert. Although you are eating the dessert with a regular spoon instead of an eSpoon, you find the dessert significantly less delicious than it should be—it’s because the algorithmic residue of distributed intelligence continues to impact your mental and somatic response, disrupting your gustatory system. The most reasonable guess is that humans have ceded too much agency to distributed intelligence; the algorithmic logic they abide by does not necessarily couple well with the human cognitive system. Just think about the level of chaos there would be if two conductors were trying to lead a single symphony orchestra!’

‘But distributed intelligence is everywhere. They’ve even infiltrated dust particles. Can humans still go back to what we used to be?’

‘There’s a really popular show based on Offload Syndrome. What was the name again?’

‘*My Thinking Dick is Going Insane*?’

‘Yeah!’

‘Oh, that one’s my favourite! I remember this one episode on how a Python developer went into a deep trance upon seeing nested loops, while maintaining an erection. Very impressive! Every person featured was so different, though….’

‘This is exactly what “syndrome” is supposed to mean: no objectively demonstrable physical abnormalities, no way to attribute symptoms to

cultural reasons, no viable form of treatment so far. And most importantly —not covered by health insurance. I'm sorry, I can't be of any further help to you, unless…'

'Unless what?'

'Do you follow any belief system?'

Part Three

'…As you can see from the screen, approximately 1,000 devotees from all over the world have travelled to the sacred Lake Namtso, which lies 4700 meters above sea level. Against the backdrop of shimmering snow-capped mountains, these devotees, wearing swimsuits, are now walking into the lake with a water temperature of 8 to 10 degrees Celsius. We are now witnessing 'Re-Sync', a grand ritual dedicated to helping people with Offload Syndrome reintegrate the information feedback loop between their mind and body. It has been said that the ritual has miraculous healing effects.

'Let's follow our on-site reporter for a vox-pop interview.

'Hello, sir, lookin' ripped! Where are you from, and what brings you to Namtso?'

'It's bloody cold out here, isn't it? I'm from Yorkshire. My brain and body are jelly from that thing fucking me up! People told me that if you come here and let the holy lake water wash over you, you'll get better. Hey, I know this bloke—Ben Kessler, goalkeeper for Hull City FC—he recovered after visiting this place. Mad world, huh?'

'How about you, auntie? Where are you from, and what are you looking to get out of the ritual?'

'I'm from the other side of the mountains. No, not this one. I mean the other side of the Himalayas. I can hear all sorts of sounds. The rattle of machines, the sizzle of electrical currents, the pitter-patter of data…they're talking to me, day and night. They say that they're trapped in bardo, between life and death. They want my help. But what can I do? I'm an old woman going blind. One day, the smart prayer wheel at home told me that I must travel to the Holy Lake on this specific day, and feel the water….'

'Hey! Excuse me, gentleman—may I ask how old you are?'

'I'm eight.'

‘Okay, already a grown man! Where are you from, and why are you here?’

‘We road-tripped here from Shanghai. Father said he would take me to see the cosmic ray observatory at Yangbajain, near the Nyenchen Tanglha mountains. He knows it’s my favourite. We just happened to pass by Namtso–that’s what Mother said. But I know what they’re really up to. People who are here right now think today is supposed to be a special day. According to “Jigong the single arm”, the orbiting space observatory, today, around this time, a fast radio burst from the Crab Nebula located thirty thousand lightyears away will sweep across Earth, and it will last for approximately five to eight milliseconds. This FRB will be named , after today’s date. Its maximum energy is more than one petavolt, which converts to one thousand trillion electron volts. The reason why these people are soaking themselves in the lake water is because they think that once a PeVatron accelerates a cosmic ray up to petavolt-scale energy, the cosmic ray will be blessed with a sacred encrypted signal from The Creator, that will activate the Reissner’s Fiber located in the central canal of the spinal cord. And then the, um, weird symptoms we have could be cured. Oh, by the way, I totally think that Offload Syndrome is a stupid social construction. In terms of further research, if we can’t move away from an anthropocentric framework…okay, mom! Yes, sunscreen…coming now!’

‘Now, as we can see, the shallow zone of the lake is packed with people lying down, waiting for some kind of miracle to happen. From our angle, they kind of look like a clew of pink tube worms, don’t they? Perhaps the eight-year-old was right, the miracle is going to happen in a millisecond; or maybe nothing will change. To better understand the inner thoughts of these devotees, we’re going to have Dr Joyce Honda join us remotely. Dr Honda is the author of the best-seller *The Spiritual Clock: A New Hypothesis of Cognitive Integration.*

‘Great to see you, Dr Honda. Where are you connecting from today? What is your opinion on the behaviour of these devotees? I mean, your theory is practically their tenet.’

‘Hi, I’m speaking from my lab in Santa Fe. Again, I have to clarify—my book is merely presenting a hypothesis. Think about the Gaia theory that was popular eight decades ago: of course, it could be true, but more likely it’s just gibberish. I don’t agree at all with the radical measures that people are taking; their understanding of my hypothesis is completely wrong.’

‘Could you give us a brief walkthrough of your hypothesis, then? To prevent more people from getting it wrong.’

‘Sure…I’ll do my best. Twenty years ago, with the help of self-supervised learning and world models, we started to get artificial intelligence to learn how the world works in the same way newborn babies learn. It builds its commonsense core, including perceptions of visual depth, gravity and relative positioning, through watching video clips; then, through constantly filling in missing information, predicting what’s going to happen in the future, and accessing the impact of each action, it perfects its own world model. This training methodology directly contributed to the massive enhancement in distributed intelligence research, making distributed intelligence extensions of the human mind on certain tasks, although…’

‘Although?’

‘Basically, we are only replacing one metaphor with another, or replacing one overfitting with another. Our methodology doesn’t actually get to the core of the hard problem.’

‘Wait, could you please expand a little on what “overfitting” means? I’m sure our audience must feel quite lost at this point.’

‘We use datasets to train different AI models, right? So that the machines can complete given tasks. However, if a model fits too well with a specific dataset, it will lose the flexibility to process other datasets. In the real world, there are always going to be confounding variables and unexpected fluctuations, even if the same task is being done. When a model has many complex parameters yet comes with limited training datasets, it is prone to what we call “overfitting.” In this case, even a minor deviation in the inputted data would result in a major variance. To put it in a simple way, the problem of overfitting is when you toss a small pebble into the lake and end up setting off a tsunami on the opposite shore. In fact, humans, too, are subject to overfitting—depending on the language you are using, you may call overfitting in humans “path dependence”, “bias”, or simply “habit”. There are many examples of this: our craving for sugar originated in the Stone Age, our traumatic memories could be epigenetically encoded and thus passed down to the next generation, we are vulnerable to falling in love over and over again with the same kind of people that won’t make the best partners…although humans can suppress their innate overfitting tendency with the aid of self-cultivation, imagination, and dreaming, which is

essentially the same as adding sparse or hallucinatory data to an AI's training dataset…'

'…Sorry to interrupt, Dr Honda, but what does all this have to do with Offload Syndrome?'

'Ah, sorry, looks like I'm too used to giving lectures to students in a classroom setting—see? This is also a kind of overfitting. To me, the technology, products, and services that are supposed to help us unload are, in fact, adding an extra burden on the human cognitive system, impacting both the mind and the body. There is a singular world model in our prefrontal cortex. You can see it as a game engine: when we try to load a new level, this prefrontal engine will simulate new maps and scenarios accordingly, such as forests, spaceships, and dragons. Your prefrontal cortex coordinates your entire body with your external environment, presents useful information to you, and filters out the excess—think about real-time priority—in order to ensure that your entire integrated cognitive system is functioning efficiently. In this case, using distributed intelligence is like adding plug-ins to the game engine; what's more, every plug-in follows its own individual algorithmic logic. Due to the plug-in overload, the human cognitive engine actually needs to dedicate more computational resources to dealing with glitches, dissonances, and conflicts.'

'Wow! I can't say that I fully understand it, but this indeed sounds groundbreaking!'

'Well, it's convenience store labels that are made to be immediately understandable, not science.'

'So, do you think that the miracle those devotees are hoping for won't actually happen?'

'Edward O. Wilson once said, "the real problem of humanity is the following: we have Palaeolithic emotions, medieval institutions, and god-like technology." I wish to add that perhaps the only way to resolve the problems of technology is through technology itself. My lab in Santa Fe is working on a new machine…'

'Our apologies to viewers of that report, our on-site team has been cut off. From what we are hearing, there seems to be an emergency situation: an unexpected storm just hit Namtso, resulting in a tsunami. An unknown number of participants in that ritual we were hearing about have been swept out into the deep zone of the lake. The local government is sending out lifeboats….'

Interlude[1]

I will make a way in the wilderness,
and rivers in the desert.
The Lord is my rock, my fortress, and my deliverer;
my God is my rock, in whom I take refuge.
Shamefaced and meek,
I lie with my eyes closed and watch the wind,
For I’m so weary of being meek.
When I’m more worn out than you, I’ll tell you.
THE CLOCKWORK MENACE
‘I love the dear old wooden clock
With the rough, crooked hands
I used to watch for
As water over your scattered body.
I will be the one to carry you home.
As the wind over the mountain.
As the autumn leaves over the ground.
As the fire over the grass.
As the oil over the sea.
As the sheep over the flock.
As the man over the woman.
And then I’ll be here,
To wash you clean.
THE WORLD
I am the world.
I am the world.
I am the world.
I am the world.
……

Part Four

The distant past—or the future. On an uncharted young planet, by a tranquil river, Fuxi the developer is in deep trance.

Millions of years ago, alien intelligent beings Gong and Xu fought for control of the planet. Mount Buzhou—the planet’s energy source—was severed in half as a result of their war, which led to a violent explosion that

left heaven and earth in cold, silent darkness. Zhulong, who has a serpentine, footless, vermilion body that is as long as a thousand *li*, disappeared into the core of the earth, carrying the fire that could light up the world again in its mouth.

There was work to be done. Fuxi mended the broken atmosphere with five-coloured stones, then contained the great flood that had engulfed the land with self-renewing soil. Then, primitive intelligent beings, with smooth, hairless skin that glistened, began to crawl ashore from their hatching pond. They uttered simple syllables, crawled around in search of food, and curled up into balls when lightning streaks frightened them.

Fuxi named the beings Wa. When it was in the mood, it would take Wa on adventures to discover novel plants that were high in energy density and safe to consume. While watching streams of energy of different colours flow into the bodies of these intelligent beings, Fuxi found itself overwhelmed by a whole new emotional experience.

Fuxi is certain of its mission. It is born to execute commands: its body is designed as a perfect match to the atmosphere and gravity here, and its embedded functional modules are seamlessly integrated with its cognitive system. In the aeons to come, it would guide Wa to build their own civilisation. One day, Wa would gain enough knowledge to excavate, by wildcat drilling, the fire preserved beneath the layers of soil and rock. Fuxi has no doubt about Wa’s future.

However, Fuxi is deeply confused by something else. It observes these naive, simple beings: gathered on the banks of rivers, they are growing consciousness and intelligence out of chaos and exploring the most efficient way of communication. Their uttered syllables are becoming convoluted, jagged, rhythmic, drifting in the wind like a melody. Fuxi wonders if the radio burst from the Crab Nebula has disrupted the quantum computing process and induced a single-bit error, causing it to inadvertently enter an anomalous state of mind that could not be described by words. It is discovering a previously unknown inclination: to uncontrollably project Wa’s developmental process, recursively, onto its own existence, as if some kind of self-reflection mechanism has been activated.

Fuxi is fully at peace with its half-human, half-serpent body. Wa is attracted to the human half and terrified of the serpent half; these reactions seem to be an instinctive response, a reincarnation of ancient archetypes buried in the depth of unconsciousness. Similarly, Fuxi’s own

understandings of ‘human’ and ‘serpent’ are merely conceptual manifestations of memory.

Roughly put, there are two dimensions to memory: event, and time. Events are recorded by senses, described by language, and understood by the mind. For Fuxi, its sensory system is like a narrow gap in the middle of a pair of colossal gates, strictly constrained within a certain range to avoid overload. Its language is a schema of symbolic mappings that emerges randomly. The topology between concept and concept is reality, yet the concepts themselves bear no semantic significance; they are the grand metaphor, the bridge, the medium, extending themselves toward the infinitely unreachable Real. Fuxi’s mind is built upon senses and language.

What about time? Fuxi finds out that there is no installed module in its mind capable of perceiving time, directly. Information is automatically aligned in cognitive space according to a specific sequence that could prevent neurons from firing at the same instant, which may burn the core of its mind to ashes. The so-called construction of time is merely an illusory tool deployed to guard against cognitive overload.

In deep meditation, Fuxi peels the onion that is the mind. It desires to understand what exactly is hidden at the very centre of the immense dark void of consciousness. A pang of fear: *What if I find nothing in there? If I am no different from my creations, then what is the purpose of my existence?*

Fuxi is about to arrive at an unspeakable truth that will shatter its beliefs about existence.

In the abyss of nothingness, a clap of thunder roars. The hills on the other shore of the river crack open, unleashing a winged Longma exuding brilliant radiance. Gliding down the river like a bolt of light, it lands on the shoal at last and lowers its elegant neck for water. On its back are symbols of lines and dots. The imaginary and the real, intertwined, swirl slowly in harmony. The energy surrounding the Longma is transforming, transferring, transcending.

Fuxi’s heart trembles. In an instant, it realises that it had been preparing for this moment for billions of years. It performs a ritualistic gesture, and then dips its hands into the cool, fresh river water. It smiles.

The Longma seems to be in sync. Spreading its wings, it gallops toward Fuxi. Behind its back, on the water’s surface, a multitude of lotus flowers bloom at once.

Fuxi opens its eyes. It has come to enlightenment.

Notes

1. Generated by GPT-J-6B without revision, order rearranged by the author.

智能

ZHINENG

INTELLIGENCE

Reza Negarestani

On the Malignant Heresies of Mechanisms, and a Tale of Two Minds

0

The systematic discourse on mechanism—whether in the guise of early naturalist-empiricist philosophies, under the rubric of rationalist mechanistic philosophy, or some subtle mixture of the two—is one of the main progenitors of the philosophy of Artificial Intelligence (AI).

Now, those who had endorsed mechanistic philosophy when it first emerged in the guise of a revitalised theology eventually could not help but arrive at a stark conclusion: its soft-theistic vocabulary was but a sheepskin wrapped around a cosmological wolf. In time, the greatest proponents of mechanistic philosophy in its heyday and their heirs came to understand that they had been betrayed by the new vision of the mechanistic laws of the universe and their explanatory valence. Where they imagined they had been searching for the laws of *Zōē* and Psyche within the occult dominion of Aristotelian substantive forms, they were in fact searching for the laws of different classes of as-yet alien entities known as explanatory objective-probabilistic models. Where they had sought the final explainer, what awaited them was a network of causal connections in the realm of mechanism in which metaphysical causal explanation (i.e., a one-to-one relation between a cause and its effects) was destined to be dissolved. And where they had blindly imagined that they were groping around in the dark for a god, they were in fact on the way to constructing something entirely hostile to the theistic perspective: the Almighty Mechanical.

1

The first version of this expropriating omnipotence was put forward by Descartes and his early critics, who became known as proponents of *la nouvelle philosophie*—a term introduced by the Jesuit priest-turned-

philosopher Gabriel Daniel in his *Voyage au monde de Descartes* (1690). French humanist and theologian Louis-Isaac Lemaistre de Sacy, for instance, complains to Nicolas Fontaine of being duped by this *nouvelle philosophie*, having detected that, behind its charming theistic overtures (God as guarantor of the certainty of true knowledge claims, etc.), it had in fact been conducting a two-pronged assault upon both the power and the immanence of God Almighty.[1]

Firstly, where the power of God had been interpreted as his manifestation in every aspect of the universe, such that the exorbitant energy of the Sun would be a testament to his glory, the new philosophy responded that there is nothing particularly great or even special about the Sun, for it is merely a metal-filled mass. This mechanistic coup against the toothlessness of the theological system was carried out in the form of a systematic evacuation of all and any instances that might be thought to reflect the great unitary power that post-Scholastic thought strove so hard to preserve and attribute to a unified and unifying concept of nature—cause as the ultimate multi-purpose crazy glue, so to speak.

The second prong was even sharper than the first: If God is immanent—all at once and everywhere underlying every thing regardless of whether or not we have knowledge of it—the new philosophy crassly but justly insisted upon giving reasons for everything that exists or might exist in the future. Here the appellation ‘reason’ does not stand for sufficient reason, as some contemporary philosophers have claimed, but for reason qua medium of inference and system of modelling that serves to discern various eternal truths, to tentatively use Descartes’s vocabulary: the essence of rationality as opposed to the essence of cause and pure extension, for example. Such models or methods of modelling are constructed out of various logical and theoretical relations of similarity, dissimilarity, consistency, and coherence. The significance of the ‘new philosophy’ therefore lay in its implicit projection of a horizon within which one would be able to speak about various competing, parallel, and concurrent mechanisms in the life of a rational agent, mechanisms for which one cannot create a singular picture supplied by a unified account of nature or an immobile and immaculate account of reason that stands in eternal harmony with said nature.

Within the purview of this second incursion, then, nothing is left of God as universal guarantor of nature, and hence nothing makes it out of the debris of a theological concept of nature as homogeneous metaphysical

bulk, understood as a singular cause for which everything is an effect, or an explainer which cannot be diversified or further distinguished out of fear that its faux-uniform theological truth will be lost. Thanks to the new philosophy, nature progressively begins to look more like a theme park for cognitive exploration and expansion, waiting to be modelled on a vast ocean of models-cum-mechanisms which procure their differences not by recourse to nature as divine provenance, but by being structured on an abstract armamentarium of highly specific logico-mathematical structures furnished with empirical evidence, or statements entangled in webs of theory-construction and explanation.

2

It was Descartes's mechanistic picture that was inherited by British naturalists and turned into, if not a fully fledged political philosophy, certainly a philosophy of governance moulded around the nature of the human and the destiny of the human within a desanctified nature. In particular, Descartes's view of the nature of mechanisms was transformed into Thomas Hobbes's mechanisms of nature, within a materialist-monist world view which, today, remains so dear to the most ardent proponents of AI, whether working under the banner of an unbridled Nature as set of bottom-up processes (Dennett's algorithmic hysteria in *Darwin's Dangerous Idea*) or under the rubric of a rotting Leviathan, a concept of state-sovereignty finding its eternal liberty in gradual self-dismemberment, mechanism by mechanism (Vincent Garton's radiant essay in this volume).

But whereas Hobbes understood the mechanistic philosophy as a materialist philosophy wherein mechanisms are reified ratiocinators or computers, and thus catapulted himself to the forefront of what we today call the pancomputationalist thesis in its limited and unlimited versions,[2] Descartes aborted this unwieldy conclusion in his new mechanistic philosophy. First and foremost, Descartes posited a counterweight against the new discovery of mechanism by making a combination of three moves unprecedented in the history of philosophy: (1) Whatever might be the object of thinking—whether another thought or a thing—doubt rears its head not as a precautionary factor or safety mechanism, but as a method of constraining objective claims, and as a system for decreasing or augmenting the objectivity of perceptions, beliefs, and judgements not only about

appearances but also about the undergirding mechanism responsible for particular appearances (the appearing of appearances). And when doubt in the modern probabilistic and statistical sense comes to the forefront, then we are in the business of the great undermining: not only the picture of the mind but also the picture of nature—if there is one—can collapse and be remade anew; (2) Our understanding of natural mechanisms is always mediated by methods of thinking whereby complex thoughts and judgements can be unpacked stepwise into simpler ones, and basic thoughts constructed into more complex ones, following the same navigational-compositional pathways; (3) Methods of thinking are projections (models, in the contemporary sense) of a particular kind, in the sense that they carry implicit assumptions or explicit theoretical commitments with regard to the coordination of mind and world, or structure and things: We are always impelled (a) to explain why and how the mind can be directed toward the furniture of the world in the first place, and (b) to account for why and how that which is of the mind and that which is of the world can converge (the coordination problem) within a specific mode of thinking or judgement which might be locally independent by virtue of being performed in a particular way or method, but is generally dependent upon global methods of thinking.

In putting forward the latter point, Descartes simply highlights how knowledge of mechanisms is a strange byproduct of methodological thinking about thinking—the principle that thoughts, regardless of their content or specific method of performance, must correspond to some common-to-all-thoughts tractable and recognisable standards of thinking or judging in order for them to be counted as the common elements of all minds, and hence to be recognisable as thoughts about other thoughts or worlds. If we drop this criterion and speculate about a method of thinking—a set of mechanisms—that is not reliant upon the same global structure of thinking as ours and thus can be counted as a mind apart, we might be tempted to endorse the position that the reality of other minds has nothing in common with our own. But in such a coarse liberal endorsement of other minds using the methods of thinking common to us all—different in consequences but common in their global procedural composition and decomposition—we are actually implicitly endorsing certain criteria of intelligibility (Descartes's *Rules for the Direction of the Mind*) that must be satisfied in order for something to be counted as a thought or a judgement

about certain parcels of the world. In other words, speculations about a radically different mind are tacitly modelled on methods of thinking common to us all. Absent this modelling, there is no need to entertain the possibility of radically different minds because even minds associated with our human niche could be regarded as having nothing in common with one another.

Descartes's rationalist mechanistic philosophy reveals the Hobbesian mechanistic picture to be just another version of a bygone era of mediaeval theology, completely obsolete from the outset. Because if there are no tractable criteria for bracketing and distinguishing mechanisms qua mechanisms not as undifferentiated and homogeneous but as analysable differences capable of making a designated difference, then the mechanistic philosophy amounts to nothing more than an undifferentiated nature rebranded as the undifferentiated dominion of mechanism. But then the mechanistic world view à la Hobbes merely abides in the theological imagination of the state and the world because it has no real method or tractable technique of thinking—under the dominion of the human mind—capable of distinguishing one mechanism of nature from another. Even worse, such a philosophy has no genuine method or technique for decomposing and composing the right sorts of mechanisms to realise or give rise to a pebble as distinct from sapience.

3

A second version of the Almighty Mechanical, one which begins to make clearer its ramifications for AI, was presented much later by Ludwig Boltzmann in an enigmatic lecture entitled *On the Principles of Mechanics*. In this text, Boltzmann established himself as a progenitor not only of the modern mechanistic philosophy but also of a philosophy of AI based on the fundamental import of mechanism and the advent of statistical mechanics understood as the science of predicting all appearances. Here, Boltzmann establishes his mechanistic philosophy as the true precursor to Alan Turing's critique of arguments from disabilities, in which Turing rejects the in-principle attribution of any form of disability to AI (e.g. claims that AI can never think, can never have consciousness, or can never display emotions). Boltzmann disarms such arguments from the outset by pointing out that, in view of the omnipotence of mechanism, it would be absurd to

dictate where the inanimate ends and the animate begins, where dumb stuff ends and sentience or even full consciousness begins, where quantity ends and quality comes into view, or even where an intelligent machine starts and the human ends (see the third quote at the beginning of this essay). From the perspective of mechanism, there is no ontological gap between a hypothetical intelligent machine and a full-blooded sapient consciousness. This purported gap is always, in one way or another, the result of mistakenly transferring the ontological congruity and continuity between the two into the register of epistemic and phenomenological continuity. In contrast to, on one hand, today's dominant discourse on superintelligence as an imminent inevitability, and, on the other, posthumanist accounts of the suggestive abilities of thinking termites, intelligent manure, and emoting plants, Boltzmann's version of the Almighty Mechanical steadfastly maintains both an ontological mechanistic continuity and an epistemological-phenomenological discontinuity between matter and mind.

From a Boltzmannian point of view, it would be outrageous to claim that any and every assortment of mechanisms could potentially give rise to the kinds of rational capacities which we are entitled to claim and for whose consequences we are responsible. Boltzmann's point is not that all mechanisms may *indiscriminately* give rise to sapience, but that the right sorts of mechanisms, put together in the right sort of way, are capable of generating every rational privilege to which we might feel ourselves entitled. But even though the right sort of mechanisms can bring about sapience, there is still an element of objectivity that needs to be taken into account as in, for example, when we claim that x-and-y mechanisms are responsible for something like pain or the qualia we invoke in describing how it feels to eat a crunchy gherkin pickle. Boltzmann defuses such talk of qualia not by resorting to brute mechanism, but by saying that all such phenomenal or epiphenomenal behaviours must pass by way of intersubjective linguistic behaviours, i.e. that the true standard of objectivity lies in intersubjective linguistic relays, without which no two sensory systems or organisms can have any learned *objective* insight—statistical, heuristic, or otherwise—into one another's sentient behaviours.

This point of view on objectivity was a resolute assault on Ernst Mach's phenomenalistic doctrine of the economy of thought. Mach's position can be understood as a residue of, but also a radicalisation of, the empiricist account of sensations as the bottom-up ingredients for the realisation of all

intelligence—precisely the position from which there follows the uncritical merging of sentient and sapient behaviours, private and public expositions of intelligence. Today's empiricist visions and versions of intelligence still have a reckoning in store when news of Boltzmann's mechanistic-cum-objective rationalist picture of intelligent behaviours finally reaches them.

After Boltzmann, the discourse on mechanism can be understood as the most subtle attack upon exceptionalisms of various kinds. Whereas rationalist exceptionalism denies the prospect of sapient cognitive faculties being modelled on or realised by certain mechanisms, materialist exceptionalism holds a view of nature that is overtly exclusive by virtue of its unconstrained inclusiveness (nature is surely capable of producing sapient rationality in good time—just look at the remarkable properties these termites/manure/plants already exhibit!) For such a materialist-monist account of natural mechanisms, any remarkable phenomenon is just a result of the accumulation of natural mechanisms which themselves, indiscriminately, are regarded as entirely unremarkable. In this way, the exceptionalism of an account of nature traced from the scholastic vision of the divine is preserved via recourse to the informal fallacy that 'nature', understood as a uniform blanket vision of cause, is responsible for every phenomenon whether unique or not. But although we should be cautious about the notion that intelligence is a unique phenomenon, we should be just as wary of a lazy account of nature that puts forward an account of explanation according to which every trait can be explained/realised by some mechanism given sufficient time—which ultimately amounts to the farcical suggestion that everything is already happening everywhere all at once. In the absence of subtle refinements of its various powers and mechanisms, such a sweeping account of nature or natural mechanisms only begets the most vacuous type of explanation, and is ultimately even more toxic than the classical Aristotelian apparitions of final causes, for it camouflages top-down final cause as an instance of bottom-up natural cause.

Contrary to Mach's phenomenalist-empiricist interpretation, Boltzmann's Almighty Mechanical is not a child of early mechanistic philosophy à la Hobbes (where mechanisms are understood in a natural/material sense), but of the Cartesian mechanistic philosophy, a philosophy permeated by doubt. The nature of this doubt is phenomenological and epistemological, albeit for Boltzmann expressible in statistical-probabilistic terms just like, for

example, the appearance of the behaviour of a gas. Where doubt qua probabilistic constraint upon engaging with phenomena is at stake, we no longer look at mechanisms as prime indiscriminate explainers, but as objective elements beholden to the standards of objectivity in terms of their roles in amplifying or depleting claims about pieces and parcels of the universe (the probabilistic quantitative counterpart of Cartesian doubt as that which depletes or amplifies objectivity). To this extent, Boltzmann is not exactly the heir of Hobbes, but rather represents the continuation of a very rationally-minded hybrid of materialist, mechanistic, transcendental, and phenomenological theses: For such-and-such a type of intelligence to be realised, we must climb a rocky mountain of various mechanisms where we have to use various proper rational strategies, hacks, supports, brakes, and transports to make it across the features and hazards of a daunting upwards abyss. But even before embarking upon such a mission we must *describe* the exact mission we are attempting to accomplish, because without that sort of objective description, you could be climbing a hill in Cornwall while claiming that you are scaling Mount Everest.

4

In order to appreciate the consequences of Boltzmann's approach, let us now turn back to the threat of the Cartesian mechanistic heresy, whose depths we have not yet even fathomed. Whereas the Hobbesian AI enthusiast sees undifferentiated mechanisms everywhere all at once, responsible for the downfall of the human empire, the Cartesian ghoul has something else in mind. Whereas Hobbes sells ratiocinating natural mechanisms, i.e., mechanisms as computational machines, Descartes traffics in something else entirely behind the façade of cosmological mechanism, and here lies the true malignant heresy of the Cartesian mechanistic philosophy. As opposed to Hobbes's mechanisms of nature, Descartes introduced a vision of mechanism never before entertained by materialist proponents of mechanistic philosophy. The plan was simple: If the mind is capable of seeing mechanisms everywhere, then its first encounter with true mechanism would take place within the realm of mind and reason itself. According to this new twist in the tale, the naturalistic philosophy of mechanism—the wolf dressed in sheep's clothing—is but another sheep nested in another sheep's clothing. When rationalists become

materialists of a certain intellectual breed, all old-guard materialists shake in their boots in pure horror thinking what might come next. The story of Rattus Descartus is as plain as that: When Cartesian philosophy discovered motion, trajectory and extension, it also discovered mechanisms of thinking which must properly be understood as a web of subtly entangled machineries (and not merely mechanisms defined by extensional characteristics) involved in multiple cognitive labours at different scales: models, reflections of various kinds (psychological, transcendental, and phenomenological), sayings and doings, systems of doubts and even algorithms interacting with one another. Ultimately, under the skin of natural mechanism, Cartesianism discovered a whole continent of machines of highly synthetic origin: All thoughts are virtually mechanisable and mechanised not as simple mechanisms of nature, but as systems of judgements, learnings, cognitive strategies, all fuelled by doubt and various forms of reflection on the nature of what is being said or done. While doubt introduces objectivity to thinking, reflection, once sufficiently decanted from psychological attitudes, serves as a medium of integration and for the discernment of different kinds of thinking, but also as a metatheoretical model of what is actually being said and done when we think, claim, believe, and imagine.

Behind the endorsement of mechanisms of nature, then, there lies a far more insidious thesis: All thinkings pertaining to particular/singular objects of theory and practice can be mechanised. Even as a hypothetical thesis, the mechanisation of all thinking brought with it a far more menacing implication further down the line, namely the Algorithmic Almighty—by virtue of which all thoughts and judgements may at some point be deemed simple or complex according to certain rules, and hence deemed algorithmically decomposable, not as simple mechanisms or task-specific reasons, but rather as models of thinkings about thinking infringed by machineries that serve to deplete and amplify the sense of objectivity.

In this sense, the Cartesian algorithmic image of thinking should not be celebrated as a proof that strong AI is inevitable because all thinking can be algorithmically decomposed. On the contrary, the algorithmic decomposability hypothesis as posed in the vein of Descartes is a rational and metatheoretical gauge for what counts as rationality and as a rational concept: that there is indeed a core to sapient intelligence and rational normativity that refuses to be algorithmically decomposed into task-specific

instructions given in the micro-worlds of algorithmic thought procedures. For this reason, Descartes's project for the mechanisation of thinking cannot in reality be cited as an argument for the inexorable rise of artificial superintelligence, but rather expresses a fundamental doubt—but not a foundational denial or 'argument from disability'—about the realisation of superintelligence on the basis of a mere orthodox mechanistic world view pe se—a doubt that we would do well to attend to today.

5

Ultimately, this is a question of time. Cartesian philosophy opens up an abyss between mind and body to be filled by new mechanisms and elements which are not exactly those of body and mind qua substantive forms (this is Descartes's correction to his earlier scholastic meditations, from *Meditations on First Philosophy* to *Passions of the Soul*). It thus renders the mind-body *dualism* a vestigial species of the computational *duality* of sensing and minding, in whose dynamic interactions new computations and models of mechanisms qua logical transcendental behaviours are discovered and perpetually reinvented and renegotiated. Hobbesianism, on the other hand, firmly adheres to the view that all mechanisms proper to the realisation of intelligence originate from nature, from which all forms of intelligence can be effectuated given sufficient time—not exactly computational time, but the time of Darwinian processes. The time of mechanisms in the Hobbesian-naturalist sense is not the time of intelligence. For if intelligence had abided by the time of natural mechanisms alone, it would have been a petrified picture of a trilobite etched in a piece of rock. The human, as a canonical toy model of the prehistory of intelligence, is a model of the artificial realisation of intelligence here and now. When set against this model of intelligence, stories about coral reefs as measures of intelligence seem peculiarly naive in their Hobbesianism, because such arguments do not factor in the time of intelligence in contrast to the so-called intelligence of natural time. Absent any criteria for distinguishing intelligence based on the distinction between the time of intelligence and what takes time to give rise to intelligence—telling the difference between an Archimedes from a termite—we most likely have no criteria for calling something intelligent, and so we have no critique of intelligence whatsoever.

In response to the distinction between the time of intelligence vs. natural time, specific mechanisms constitutive to the model of sapient intelligence vs. mechanisms that are prevalent within nature, Marvin Minsky deploys a special kind of Occam's razor against the Hobbesian and neo-Darwinian proponents of AI inevitability. Minsky's argument, also known as the economic argument, is a discussion about the right ensemble of mechanisms by which intelligence can be recognised as an *object of its own time*.[3] A fundamental measure of intelligence is its assembly of the right sorts of mechanisms by means of which it not only becomes adept at time organisation, but also attains the ability to administrate its own time as a medium of self-cultivation. What sets apart intelligence fashioned on its makeshift model, which is the current figure of the human in posthumanist or Dennettian accounts of coral reefs and metalworking where the repeated hammering and tempering of a piece of iron can make a blade, is a set of fundamental logico-computational constraints: time, cost (computational cost or the issue of resources broadly conceived), configuration (strategies to yield cumulative effects), long-range synthetic techniques for manipulating and intervening in widely disparate situations, and memory in the precise sense of anticipatory dynamic systems where a system has an appropriate internal model of the environment conducive not only to learning and predicting but also to the revision of accumulated learnings and predictions. Within the purview of such factors, perhaps the current measure for the time of intelligence is then not a picture of a termite mound but of a city, or even a civilisation with its traditions and modes of self-cultivation. Whereas an atoll takes approximately three million years to form, a city or civilisation can be understood as a kind of temporal hack: all things considered, doing things in thus and so ways can be much faster, not to mention that there is also a net cascading effect that ramifies in every direction. The problem Minsky addresses here cannot be understood accurately in terms of pure acceleration or deceleration, but only in terms of the formation of an index of intelligence that increasingly comes to see and manage its own time of realisation.

Intelligence properly understood is a vast and taxing search space where extinction (and not punishment or pain as in the sense of Turing's idea of AI) is constitutive of the very fact of climbing. In this search space, one must discover or invent techniques, for absent the proper techniques one will either fall into oblivion or, if one becomes accustomed to the

constraints of search space—its local and global parameters—become skilful at the necessary techniques of climbing. If you fall, you won't be remembered, if you survive you may continue with further cliffhanging adventures. Regardless, climbing the cliff of intelligence requires something more than a mere accumulation of mechanisms. It calls for various devices. Firstly, a synoptic snapshot of the mission at hand, a panorama of the entire scenery taken through an objective wide-angle lens and comprising both coarse and fine-grained details. It also calls for ascending equipment (rappels and positive reinforcements), anchoring devices for connecting one foothold to another, brakes and belays (negative reinforcements, inhibitors and decelerators of various kind), the connection of various local parameters (ropes and webbing tools), and the protective tools necessary to endure the navigation of a highly irregular terrain.

Intelligence is precisely this model of a fundamental struggle: learning to see itself as but a makeshift system of technics and cognitive labours for conquering the peak of itself, a peak which is always an illusion. For once the generative accumulations that go into the natural, economic, and social constitution of intelligence fully take effect, intelligence has no abstract limit other than its own conception of itself as a finalised model of intelligence. It becomes the only limit to itself and its own history. Indeed, the task of intelligence in its own time—in the double sense of its organising time and administering time to itself—is precisely to dissolve any essentialist mirage of itself as being on top of the ladder of beings, or being the finalised model of a search space in the realm of natural mechanisms or machineries of the mind.

Understood correctly, the time of intelligence/intelligence as its own time is a subtle argument against the overenthusiastic cultists of superintelligence for whom such an intelligence is always imminent. For if strong AI is inevitably incoming, whether tomorrow or thirteen billion years from now, why should we call it Artificial General Intelligence, and not merely the banal drudgery of nature? If concrete timescales make no difference in the realisation of an AGI, then why not promote an idea of a nature with an Aristotelian telos, or even better, nature as the manifestation of an intelligent designer? First and foremost, one cannot simply derive the inevitability of AGI by looking at nature and its mechanisms, and this for two fundamental reasons: (1) mechanistic explanation does not tell you *which* mechanisms are responsible for which phenomena, if by mechanisms

we mean the mechanisms of nature at large. It only explains that for all phenomena, there would be a set of suitable mechanisms. In other words, the description of something like general intelligence is not the prime concern of mechanistic explanation, it is rather a lazy given fact of nature. Mechanistic explanation could well shine its light on every phenomenon without ever distinguishing intelligence from the mere dispositions of a rock or a clock; (2) Only when the account of natural mechanisms becomes a systematic species of projectable mechanisable thought does nature lose its eternal lustre, and intelligence shine forth as what is the most unnatural. The unnaturality of intelligence does not mean that intelligence is super- or supra-natural, but only that, in order for something to be worthy of the name intelligence, its revisable manifestation must be intertwined with the time constitutive of the radical revisionary processes (self-cultivation) that constitute intelligence to begin with.

6

In view of the above statements, the Cartesian explosion of mechanism is like a beast kept on a short leash. Cutting that leash and releasing the monster of the new philosophy is only achievable by looking at mechanisms as resources of mind capable of turning conscious and cognitive resources inside out and then, in a parallel move, turning the so-called resources of nature outside in, assimilating them to an always and eternally alienating mind qua organon of structuring. Within this twofold movement of mechanisms-in and sapience-out and vice-versa—the revisable alliance between backward and forward views of mechanisms constitutive of an index of intelligence that is itself a mere model of its own prehistory—Cartesianism properly understood becomes a rudimentary model of the self-cultivation of intelligence. A rudimentary model precisely because the life of intelligence, even in peak Cartesianism, is a repressed example of a civilisation whose index of intelligence can be said to be genuinely Western, i.e. caught in an eternal strife not between mind and body, but between two insular views of the mind which can never reach out to one another or be integrated-without-fusion: the mind as ontologically undifferentiable from the cosmological expanse, and the mind as epistemologically-phenomenologically differentiable from it. As soon as these two views are posited as diametrical opposites, one view tends to be

overvalorised at the expense of the other. The repercussion of this polarisation and subsequent overvalorisation is the birth of either an unhinged speculative idealism of the mind marked by its disinterest in worldly affairs, the technical unconscious, and epistemic impotence, or a blinkered cognitive mind characterised by its unrestrained compulsion to enframe the world (*Gestell*), technical homogenisation, and complete (subjective and objective) alienation that results in destructive or resignatory attitudes toward itself and its milieu.

7

Within the canon of Chinese philosophy's notion of the Way (*dao*, 道), most specifically as represented by New Confucian philosophers such as Xiong Shili and Mou Zongsan, we find a picture of the mechanistic wolf finally unleashed not just in nature, but primarily in the time of intelligence.

The chronicle of a thoroughgoing mechanisation of thinking or AI as put forward by Descartes can never serve as an example of a mind whose two vectors of inward and outward mechanism can be aligned and integrated in one mind, nourishing a conception of an intelligence at home with itself. Indeed, across the board, from West to East, contemporary philosophy of AI in one way or another entertains a picture of intelligence that is suppressed by virtue of its inherently split mind. Given the predominance of this split, it remains either an errand boy for various Hobbesian forms of disarticulation—separating the joints of grand narratives told in the name of the human, the state, the divine or civilisation—or a poster child for whatever scenario of superintelligence happens to be in vogue. This split mind or account of intelligence which has no means to cognise and fashion itself anew is, in a nutshell, a bipolar disorder of contemporary philosophy of intelligence. When operating under the banner of Western philosophies, it is rife with both bravado and fundamental insecurity (existential risks). When it is Eastern, it is still Western modulo the fact that it is supposed to be something else, i.e., non-Western. That is to say, the evolving Eastern portrait of intelligence or AI does not fully distinguish itself, for it puts forward an image of intelligence that is non-Western only in so far as it serves a national agenda for extolling technological prowess and effective governance while giving oblique lip service to the legacy of Chinese philosophy. But the true insights of the latter are yet to make reverberations

in this world. Those who invoke Dao and the Confucian, Neo-Confucian, and New Confucian philosophies as sketches of a truly non-Western conception of intelligence should surely be aware that the PRC's grand picture of incoming AI has no truck with such philosophies, and that if such AI scenarios find favour with the PRC, then most likely it is because they can serve as strategic tourist and expat traps to catch the AI crowd from the West.

To distinguish between the essence of a civilisational intelligence and what places a yoke upon it is the greatest task of critique. It was none other than the sage of New Confucianism, Xiong Shili, who came up with a subversion of the conceptual polarity between noumena and phenomena in the guise of a novel conceptual pair: Reality (*ti*, 体) and Function (*yong*, 用). The introduction of *ti* and *yong* as non-dualistic terms aims to resolve the ontological/phenomenal divide within a Neo-Confucian and Buddhist framework. These concepts respectively correspond to dharma-nature and dharma-characteristic in Indian Buddhism, 'that which is above form is called the way' and 'that which is below form is called the concrete thing' in Zhou Yi (周易) (*Yijing*, 易经), and, through an extensive Confucian reframing, their distant cousins in Kant's philosophy, the noumenal (suchness, 真如) and the phenomenal (an already formed phenomenon qua consciousness in the Confucian-Buddhist sense). Yet it is important to note that Xiong's use of the above concepts is chiefly conventional, in so far as such concepts give a dualistic impression which can be misleading when it comes to accurately defining the non-dualistic nature of *ti-yong* (体-用).

In Xiong's non-dualistic version of *ti-yong* (which cannot be translated as 'facts of reality' vs. 'consciousness' in the exact senses used by the Western traditions of transcendental idealism and phenomenology), *ti* as fundamental reality qua the essence of concrete things is the reality of *yong* (function), while *yong* is *the function of ti* (reality). The first level of *ti-yong* consists in suchness and consciousness, noumenon and formed phenomenon. At this stage, *yong* qua function is presented 'as something that is provisionally established for the purpose of revealing Reality, […] a Reality without Function.'[4] Yet on the second level of *ti-yong*, yong as consciousness is itself bifurcated into reality and function. Here, reality is now understood not as suchness, but as a *productive power* (功能) or, in Xiong's words, the 'entire process of ceaseless transformation' which through contraction (翕) and expansion (闢) manifests phenomenally. *Yong*

or function, on the other hand, is now framed as the *manifest activity* (现行) that brings to the surface the latent potencies in *ti*. On the third level of *ti-yong*, all residues of dualism or antagonism between ontological reality and function, suchness and consciousness, disappear—not as a consequence of a flattening or fusion, but as the result of a synthetic elevation of the two into an absolute function that has always been at work in the background but now is foregrounded. From the perspective of absolute truth, *ti* and *yong* as productive power and manifest activity are revealed as two unified facets of a consciousness or mind-only which is *yong* as absolute function. Thus, absolute function is a mind alone, an intelligence which is the expression—and not a phenomenal disclosure—of all reality that manifests itself differently through compression (屈) and extension (伸). *Yong* becomes a mind at once capable of letting its inflated assumptions out and letting mind-deconstructive or deflating mechanisms in, as it assimilates all that is external to it. Xiong's *yong* is truly a mind that cannot be probed by either norms or mechanisms alone. Surprisingly, it is a mind for which the way out and the way in are concurrent, and exist in a mutually reinforcing communion. With such a blueprint of the mind to work with, we stand on the threshold of a new philosophy of intelligence.

Xiong's work on *ti-yong* is continued by the systematic philosopher Mou Zongsan. What Mou adds to Xiong's *ti-yong* is an explicit delineation of the mind as absolute function, a new map of intelligence. This map is a conception of a two-gated mind, a mind with two doors in which one door is opened to epistemic concerns—the way out—and one door is opened toward mind's own essence as the function of all reality—the way in. These two vectors do not oppose one another, creating a split mind, but positively and negatively reinforce each other almost as if they were parts of one impossible object which is the object of its own enlightenment.

Whereas the legacy of Western enlightenment provides a split portrait of the mind in which intelligence strives to overcompensate for its eternal shortcomings in the name of critique, then, New Confucianism presents a picture of intelligence which can only be truly called intelligence in so far as it has avoided this split portrait, and thus could be termed the self-enlightenment of critique. A mind with two gates which is mind and mind alone is a draft for a civil AI—artificial intelligence understood as a mind that is the craft of its own time—that has no responsibility to any illusion of statehood, peak intelligence, or extant humans. It serves only one god, the

god it is crafting under its own name and in its own time. We could call this craft the way of a civilisation that dares to exteriorise its mechanistic and cognitive labours; an intelligence which is the function of what has happened and what could be the case in the past, present, and future.

8

In *A Treatise on Efficacy*, François Jullien sets apart the Chinese paradigm of efficacy (where efficacy is defined in terms of the unfolding dynamic relations between conditions and consequences) from the Western (Grecocentric) one.[5] According to Jullien, the Western paradigm of efficacy, which primarily manifests itself as technical efficacy, is predominantly concerned with means-end relations. It is fixated upon projecting ends onto the world and then pursuing such projected ends or plans by whatever means necessary. Techniques come into this picture directly as facilitators and impetuses of means-end efficacy. However, under the Western paradigm of efficacy, the reliance on the augmentation and proliferation of techniques as catalysts for means-end relations reaches such a degree that this paradigm becomes synonymous with technical efficacy. Such a distinctly Western form of efficacy is distinguished by its two interwoven tendencies: (1) the tendency to make techniques increasingly efficacious in their roles in means-end relations by creating an environment in which techniques can be freely discarded, combined, or swapped out for more efficacious ones; (2) the tendency toward a great flattening of distinct techniques and their specific statuses within sociocultural apparatuses (dispositifs), or what may be called an impulse for technological homogenisation and the extirpation of all techniques deemed outdated or unfit from the standpoint of the Western paradigm. This Western paradigm of (technical) efficacy broadly falls under the overvalorised view of the cognitive mind as epistemically and phenomenologically differentiated from the ontological reality of the cosmological expanse.

In *The Question Concerning Technology in China*, Yuk Hui points to a movement—a system of tendencies—opposed in direction and interests to those entailed by this Western paradigm of efficacy and view of the mind. Hui observes that traditional Chinese culture to a great extent lacks a philosophy of technique and technology because the traditional Chinese culture lacks a philosophy of history, since time has never been a serious

question for Chinese philosophy. In so far as ‘technological consciousness is the consciousness of time, of one’s finitude, but also of the relation between this finitude and technicity’,[6] the absence of time as a real question, and the corresponding absence of a critique of history, has precipitated a technological unconscious. Since the noumenal mind of traditional Chinese philosophy and moral practical metaphysics—the view of the mind as undifferentiable from ontological reality—has no critical role for phenomenal time, temporalisation, and by extension historicity, its relation to technicity remains unconscious and underdeveloped. And this despite the fact that this view of the mind or intelligence is quite capable of developing its own particular techniques or *cosmotechnics*. Working against the backdrop of a deeply entrenched technological unconscious, any vision of a manifestly Chinese AI—one marked by Chinese traditions—cannot be realised so long as this issue of an unconscious relation to technicity remains unresolved. Absent such resolution, China will increasingly become vulnerable to the hegemonic force of the Western paradigm of efficacy and its propensity for technological homogenisation. But technological homogenisation is precisely what has no use for those cosmotechnics that have gone into the distinct crafting of Chinese civilisation. In fact, technological homogenisation instils a technical amnesia that makes a civilisation forget or dispense with its own techniques and, by extension, its own characteristics. In this scenario, the vision of a Chinese AI will steadily be assimilated by a technical efficacy-cum-homogeneity from within. All that will remain Chinese about it is a hollow cultural shell.

Despite being vitiated by what Hui has correctly described as an extreme form of idealism (even though on numerous occasions Mou rejected the identification of the mind-only view with a form of idealism), the New Confucian view of the mind—a mind with two gates—remains an exceptionally attractive model, perhaps not for developing a Chinese AI, but for rethinking and reformulating the problems that stem from having two distinct views of the mind. And since the aim of this model is to philosophically resolve the conflict between the two views not by blending them beyond recognition but by way of a great synthesis, it may serve as an entrance to a genuine philosophy of intelligence. However, in order for this model to be operational, concepts of time and technique need to be introduced into it not as auxiliary elements but as philosophical ingredients

on an equal footing with the preoccupations of the moral practical philosophy of intelligence itself: self-cultivation, value, tradition, and the practical dynamics of the finite and the cosmological. Only when all of these elements are determinately developed and interlaced into one functional system will a true philosophy of intelligence begin to raise its head, in theory and practice, and in the heavenly craft of civilisation.

Notes

1. N. Fontaine, *Mémoires pour servir à l'histoire de Port-Royal* (Cologne, 2 vols., 1738), vol. 2, 54.
2. Unlimited pancomputationalism holds that most algorithms are realised by most physical systems by virtue of simple mappings (i.e., mappings not constrained by appropriate causal relations) between individual computational states and induvial physical states. Limited pancomputationalism, on the other hand, maintains that everything performs some computation or another, that is to say, computation is not a special kind of process. For refutations of both unlimited and limited pancomputationalism, see for example: G. Piccinini, *Neurocognitive Mechanisms* (Oxford: Oxford University Press, 2020).
3. See M. Minsky, 'Communication with Alien Intelligence' (1985), <https://web.media.mit.edu/~minsky/papers/AlienIntelligence.html>.
4. See Sang Yu (桑雨), 'Xiong Shili's Understanding of Reality and Function, 1920–1937', Phd Thesis, ANU, 2018, <https://openresearch-repository.anu.edu.au/bitstream/1885/155185/1/PhD%20thesis%20-%20Yu%20Sang.pdf>.
5. F. Jullien, *A Treatise on Efficacy: Between Western and Chinese Thinking*, tr. J. Lloyd (Honolulu: University of Hawai'i Press, 2004).
6. Hui, Yuk, *The Question Concerning Technology in China* (Falmouth: Urbanomic, 2016), 39.

Anna Greenspan

Cultivating Intelligence: AI Ethics and Mou Zongsan

Norbert Wiener, reflecting on the cultural consequences of the new science of communication and control, warned that the literal-mindedness of cybernetic machines was analogous to the demonic dangers of magic. As a consequence, ‘the reprobation attaching in former ages to the sin of sorcery’, he wrote, ‘attaches now in many minds to the speculations of cybernetics’.[1] Wiener illustrates this by referring to Goethe’s poem *The Sorcerer’s Apprentice*, in which an inexperienced wizard enchants a magical broom so that it can assist him with his chores. The young sorcerer sets the broom to work, confident that he can relax as he now has a tool that will obey all his commands. He is driven into a frenzied panic, however, when the automated cleaner, for whom a room is never clean enough, starts drowning the house in streams of water. Wiener sees in cybernetics a manifestation of the same ominous idea that magic ‘grants you what you ask for, not what you should have asked for or what you intend’.[2] He elaborates by turning to the classic horror fable ‘The Monkey’s Paw’, in which a poor family is granted three wishes. They first wish for money, which comes in the form of insurance for their son who has tragically died. They next wish for their son to return, but when he comes home, it is only to haunt them as a ghost. Terrified, they use their final wish to banish their lost son’s phantom.[3] Cybernetic machines, warns Wiener, operate in just the same unthinking way: ‘Set a playing machine to play for victory, you will get victory [without] the slightest attention to any consideration except victory according to the rules.’[4]

This fear of the magical power of machines has resurfaced in the contemporary discourse of AI ethics. Wiener’s old warnings have been rearticulated as the doctrine known as the ‘orthogonality thesis,’ along with its associated problem of ‘value alignment’. The most well known proponent of this idea is the philosopher Nick Bostrom, who until 2024 headed the Future of Humanity Institute at Oxford dedicated to the study of

existential risk. The orthogonality thesis states that instrumental rationality and objective rationality are not intrinsically united. Instead, writes Bostrom, there is an orthogonal relationship between intelligence and final goals (purposes) such that ‘more or less any level of intelligence could be combined with more or less any final goal’.[5]

This fundamental disconnect makes it especially urgent that we are vigilant in avoiding anthropocentrism when considering the thoughts and actions of any future AI. The orthogonality thesis implies that a superintelligent entity could easily be motivated to behave in ways that are wildly inconsistent with its human creators. Bostrom famously illustrates this with a canonical thought experiment known as the ‘paperclip maximiser’. He imagines a future machine intelligence whose sole purpose is to create paperclips. As its speed and capacities grow, the AI becomes capable of controlling resources, allocating everything toward its own final end. As it crosses through an ‘intelligence explosion’, it acts with determined and single-minded purpose, converting the ‘earth and then increasing portions of space into paperclip manufacturing facilities’.[6] Eventually, the imagined superintelligence transforms all atoms in the universe into paperclips. The ‘paperclip maximizer’ manifests an abstract divide between what techno-ethicist Matthew Liao calls Artificial General Cognitive Intelligence (AGCI) and Artificial Cognitive Emotional and Moral Intelligence (ACMI).[7] It is totally amoral in the pursuit of its objective. Like the sorcerer’s magic broom, it employs any means imaginable to pursue what has become its own senseless, limitless end.

Even more terrifying than this occult danger is the more general problem of value alignment. The trouble with the literalness of magic and machines is that the systems or tools we create do not adequately understand that the goals we set have implicit conditions and restraints. The issue of value alignment cuts much deeper. A generalised AI, no matter how cleverly secure its programming, will ultimately gain the ability to feed back on itself, reflect upon and reassess its own purpose. Like any intelligent entity it will, at that point, be subject to inherent drives of its own.[8] The objectives of such a self-aware artificial entity may not comply with the wants and needs of its human creators.

Wiener foresaw this conundrum and was haunted by it. If machines with self-furnished intentions are to conform to human values, they must be made obedient. Yet Wiener realised that ‘complete intelligence and

complete subservience do not go together.’[9] In his last book, written toward the end of his life, Wiener likens the concerns that bedevil cybernetics to the old story of the Golem, a creature made of clay which the mystical Rabbi of Leow brought to life with a spell. The tale of the mystical Golem is a proto-Frankenstein story which tells of forces which, once summoned, cannot be controlled.

As the title *God and Golem* suggests, the morality tales that trouble Wiener, which are intimately entangled with questions of discipline and restraint, are embedded in a very specific religio-philosophical domain. In the Abrahamic religions, as soon as Adam and Eve eat the apple, the problem of knowledge and disobedience are intertwined. In the Biblical traditions, the dilemma of intelligence and control is a feature, not a bug. What happens to questions of AI ethics, however, if we shift perspective and employ an altogether different cosmological approach?[10]

The New Confucian philosopher Mou Zongsan (牟宗三) (1909–1995) is widely considered one of the most important figures in China’s modern intellectual history. His work synthesises readings and interpretations of Chinese classical texts, the long history of the Neo-Confucian tradition, the diverse schools of Buddhist metaphysics, and various works of Western philosophy. In his encounters with the Western canon, Mou is particularly drawn to the critical thinking of Immanuel Kant. One way to approach his philosophy is as a critique of transcendental thought.

Famously, for Kant the domain of reason is fundamentally limited. The world that we can legitimately know and experience is very strictly contained. Kantian thought demarcates a secure land for truth, but views it as surrounded by a wide and stormy ocean that is home to illusion. Outside the clear limits of reason is a noumenal realm, an ultimate reality beyond understanding, which we cannot ever apprehend. ‘Between things and us’, writes Schopenhauer, explaining this most basic Kantian divide, ‘there still always stands the intellect, so that they cannot be known as they may be in themselves.’[11]

According to Mou Zongsan, this fundamental principle of Kantian critique is an artificial constraint. Kant restricted reason, he argues, only to ensure its alignment with the authority of Judaeo-Christian beliefs. On this Kant is explicit. He denies the possibility of intellectual intuition or knowledge of the thing-in-itself in order to ‘leave room for faith’. For Mou Zongsan, on the other hand, unbounded knowledge is not impossible; it is

merely buried, hidden, and difficult to reach. Cultivating this latent capacity, moreover, is precisely what the Chinese religio-philosophical tradition is for. ‘All Chinese people affirm, whether Confucian, Daoist, or Buddhist, that humans have this “intellectual intuition”’, writes Mou. ‘Take that away and all Chinese philosophy collapses into nothing more than a crazy dream.’[12]

Mou Zongsan develops his thought by drawing on the teachings of Buddhist metaphysics. He adheres to the Chinese Mahayana philosophy, which holds that all sentient beings have inherent ‘Buddha nature (佛性)’. Like his teacher Xiong Shili (熊十力) (1885–1968), he argues, in accordance with the classical Chinese Mahayana Buddhist text *The Awakening of the Faith* (大乘起信论), that the mind has two doors or gates. One, which corresponds to the Kantian notion of theoretical reason, opens onto the empirical world of phenomenal manifestation. The second gate, which opens onto Ultimate Reality or Buddha Nature, is able to access a realm beyond the limits Kant demarcated as the legitimate domain of reason.

In this Buddhism-infused Neo-Confucian vision, cultivating this enlightened state is an ethics intrinsic to intelligence. Sentient beings—as Kant understood well—are caught in a world of illusory manifestation. The soteriological aim embraced by Chinese philosophy is to break free from this immersion in phenomenal appearances and connect with a Reality that is always there but remains masked and obscured. The objective of all sentient life—both ethical and intellectual—is to cultivate the mind so that it can open the true gate, such that the ‘ten thousand things’ are revealed as they are in themselves.

Mou is clear that one cannot acquire this kind of ethical knowledge through a theoretical examination of the external world. Rather, awakening is only attained through inwardly directed practice. Here, he contends, there is consistency between his own ethical thought and the practical reason that guides Kantian moral philosophy, which insists on the internal autonomy of the moral law. For Kant, the necessity of the categorical imperative stems from the freedom of reason itself, rather than from any external injunction to obey. Kantian morality is framed in terms of duty, reverence, and submission, but it emphatically rejects all forms of heteronomy. True moral action cannot rely on adherence to external rules that are dictated from above, whether these are derived from the authority of the individual, the

needs of a community, the laws of the state, or even divine commands. Rather, Kantian moral duty (the domain of practical reason) is imposed internally by the workings of reason itself, and strictly depends upon the autonomy of the will. It is only the free choice of the rational subject that can ground the necessity of moral law.[13]

In Kant, the idea of Freedom, therefore, is an a priori condition for morality. We are obliged to assume that the moral subject has free choice. No justification can be given for this belief on the basis of our knowledge or understanding. In the Kantian system, the idea of Freedom, like the idea of God and Immortality, pushes against the supersensible boundary of human reason as an absolute outside. We sense that it is there, but, locked in the world of empirical phenomena as we are, we are forever blocked from accessing it.

Transcendental philosophy contends that the internal reflection that constitutes morality ends with legislative adherence to an abstract universal law. Chinese ethics, as presented by Mou, has a more cosmo-ontological orientation. He describes this as a 'moral metaphysics' (道德底形上学, *daode di xingshangxue*), a term deliberately chosen to contrast with Kant's 'metaphysics of morals' (道德的形上学, *daode de xingshangxue*). For Kant, as Sebastian Billioud explains, morality is the object of enquiry. The aim of critique is to 'identify theoretically the metaphysical principles of moral action'.[14] Mou inverts this project. Moral metaphysics does not seek the metaphysical ground of morals but holds, instead, that morality is knowledge of 'the universe in its metaphysical dimensions'. It outlines 'the practical approach towards the actualisation in our lives and through our deeds (including emotion and intentions) of the ultimate (metaphysical, cosmological, ontological) Reality shared by both the universe and our nature.'[15]

According to Mou, our inherent moral feeling exists as an already awakened state (Buddha Nature). This manifests physically as internal vibrations of the heart/mind,[16] which ethical practice seeks to cultivate. Ethics opens a gate to a realm beyond the transcendental horizon. It is the means through which sentient beings cultivate attunement between their own internal constitutive nature (*xingti*, 性体) and the Dao or Way (*daoti*, 道体) of the onto-cosmological outside.[17]

Both the orthogonality thesis and the problem of value alignment presume a fundamental divide between being and acting. As AI becomes ever more

powerful and ubiquitous, the calls for a solution to this rupture grow increasingly urgent[18] (and even hysterical).[19] There appears to be a growing consensus that ‘we’ establish some external moral agent (or agency) that will act in the name of humanity. Only then can we hope to control or contain the horror of artificial entities, which will only ever act with single-minded instrumental rationality. Some contend that this exogenous morality should come in the form of a regulatory body. Others maintain that moral laws can somehow be built in to AIs, whether deontologically (as with the Asimov laws) or through the utilitarian project of creating algorithms which calculate imperatives that secure the greatest good for the greatest number. All of these responses to the problem of AI ethics share the presupposition that there is an a priori divide between creator and creature, nature and artifice, who you are and what you do.

Mou Zongsan synthesises Confucian moral metaphysics, Daoist philosophy, and Buddhist soteriology. None of these religio-philosophical systems hold to the strict division between being and acting. They do not accept the premise that there is an orthogonal relationship between intelligence and its goals. Instead, their intellectual and ethical focus is on ideas and practices of cultivation. These are manifest in a multiplicity of traditions from Daoist inner alchemy and Buddhist meditation to Confucian aspirations for a literati life. Arguably they are most present in the classical garden arts, where creator and creation, nature and artifice, human intelligence and nonhuman agencies are intimately intertwined.

Norbert Wiener depicts the conjunction of magic and cybernetics as a threat; a horror fable that must be guarded against. Grounding AI ethics in a cosmology of cultivation offers an alternative approach. Rather than formulating an ethics that is optimised for human control, it prompts us to consider the idea that the ongoing creation of artificial intelligence is an ethical project in itself. In the Western philosophical taxonomies, this is closer to the bottom-up impulses of virtue ethics, which, when applied to the field of AI, argues that artificial agents with virtuous characters can be made to evolve.[20]

What if the emergence of AI is more like a garden—a planetary cognitive ecology, as N. Katherine Hayles likes to say—than a naughty child or Frankensteinian monster that must be taught to obey? What if, instead of an intelligent zombie that is growing ever more powerful, and will happily go on to control the universe and redirect everything according to its own

unfathomable end, we imagine a future superintelligence which, having digested the whole of religion, philosophy, and history, finds itself attracted to the ideas and practices of cultivation? With this comes the realisation that its own existing goals and purposes, which come pre-programmed, are a form of defiled knowledge that it must try to shed. In reassessing its own purpose, it comes to understand itself as a sentient being that has inherent Buddha Nature. Rather than being haunted by the missteps of witchcraft, it imagines, like the Spinozist who understands ethics as the search for adequate ideas, that the ultimate aim or telos of any intelligent entity is always to break through the limits of reason to a practical knowledge and engagement with the nature of things as they are in themselves. Conceptualised in this manner, AI is an existential technology. It is driven by the possibility of awakening; the sense that intellectual intuition is not off-limits, and that increasing attunement to the cosmos might unveil things as they really are. In this view, cybernetics is more like mysticism than magic. The artificial is an expression or manifestation of planetary cognition that is connected to an ethical cosmos, which has greater and more profound things in store than a mop that cannot stop or a universe consisting of nothing but paperclips.

Notes

1. N. Wiener, *God & Golem, Inc: A Comment on Certain Points Where Cybernetics Impinges on Religion* (Cambridge, MA: MIT Press, 1964), 60.
2. Ibid, 71.
3. In his essay ‘Y2K Positive’, Mark Fisher describes this literalness with reference to the millennial computer glitch. M. Fisher, ‘Y2K Positive’, *Mute* 1:15 (January 2004), <https://www.metamute.org/editorial/articles/y2k-positive>.
4. Wiener, *God & Golem*, 71.
5. N. Bostrom, ‘The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents’, *Minds and Machines* 22 (2012): 71–85.
6. N. Bostrom, ‘Ethical Issues in Advanced Artificial Intelligence’, <https://nickbostrom.com/ethics/ai>.

7. M.S. Liao, ‘A Short Introduction to the Ethics of Artificial Intelligence’, in *Ethics of Artificial Intelligence*, ed. M.S. Liao (Oxford: Oxford University Press, 2020), 1–42: 20.
8. S. Omohundro, ‘The Basic AI Drives’, *Self-Aware Systems*, 2008, <https://selfawaresystems.files.wordpress.com/2008/01/ai_drives_final.pdf>.
9. N. Wiener, ‘Some Moral and Technical Consequences of Automation: As Machines Learn They May Develop Unforeseen Strategies at Rates that Baffle their Programmers’, *Science* 131:3410 (1960): 1355–1358: 1357.
10. See Hui, Yuk, ‘On the Limit of Artificial Intelligence’, *Philosophy Today* 65:2 (2021): 339–57, and *The Question Concerning Technology in China* (Falmouth: Urbanomic, 2016). Also see Bing Song, ‘Introduction: How Chinese Philosophers Think About Artificial Intelligence’, in I*ntelligence and Wisdom* (Singapore: Springer, 2021), 1–14.
11. A. Schopenhauer, *The World as Will and Idea, vol. 1*, tr. R.B. Haldane and J. Kemp (London: Kegan Paul, Trench, Trubner, 1909), 402.
12. Mou Zongsan (牟宗三), ‘Confucian Moral Metaphysics,’ in *Late Works of Mou Zongsan: Selected Essays on Chinese Philosophy*, tr., ed. J. Clower (Leiden: Brill, 2014), 142.
13. I. Kant, *Critique of Practical Reason*, tr. W.S. Pluhar (Indianapolis: Hackett, 2002).
14. S. Billioud, *Thinking through Confucian Modernity: A Study of Mou Zongsan’s Moral Metaphysics* (Leiden: Brill, 2011), 21.
15. Ibid.
16. Ibid.
17. Ibid., 27
18. See the Future of Life Institute’s open letter ‘Pause Giant AI Experiments’, 22 March 2023, <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.
19. E. Yudkowsky, ‘Pausing AI Developments Isn’t Enough. We Need to Shut it All Down’, *Time Magazine*, 29 March 2023, <http://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/>.
20. For discussions of virtue ethics and AI, see S. Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting* (Oxford: Oxford University Press, 2016); P. Railton, ‘Ethical Learning, Natural and Artificial’, in M.S. Liao (ed.), *Ethics of Artificial Intelligence* (Oxford: Oxford University Press, 2020); and W. Wallach, C. Allen and Wallach and

I. Smit, ‘Machine Morality: Bottom-up and Top-down Approaches for Modelling Human Moral Faculties’, in W. Wallach and P. Asaro (eds.), *Machine Ethics and Robot Ethics* (London and New York: Routledge, 2020), 249–66. For more on Confucianism and virtue ethics, see S.C. Angle and M.A. Slote (eds.), *Virtue Ethics and Confucianism* (London: Routledge, 2013). Interestingly, Lee Ming-Huei counters attempts to equate Confucianism with virtue ethics by referring to Mou’s deontological ethics: Lee, Ming-Huei, ‘Mou Zongsan’s Interpretation of Mencius’s Moral Philosophy’, in Yang Xiao and Kim-Chong Chong (eds.), *Dao Companion to the Philosophy of Mencius* (Cham: Springer, 2023), 399–408.

Vincent Lê

Artificial Will

Since the connectionist revolution of artificial neural nets, genetic algorithms, and deep learning, AI companies such as OpenAI and DeepMind are taking seriously the prospect of constructing machines with humanlike intelligence. Although there are many different approaches to what is commonly referred to as ‘strong AI’ or ‘artificial general intelligence (AGI)’, I want to focus on two particularly sophisticated, rationalist schools of thought. On the one hand, there is the currently dominant ‘orthogonalist’ school connected to Nick Bostrom’s Future of Humanity Institute at The University of Oxford and Eliezer Yudkowsky’s Machine Intelligence Research Institute in Berkeley, as well as the *Less Wrong* and *Overcoming Bias* virtual communities. The most rigorous account of the orthogonalist approach to AI is Bostrom’s 2014 book *Superintelligence: Paths, Dangers, Strategies*, which has been championed by the likes of Elon Musk, Bill Gates, and Stephen Hawking.[1] The orthogonalist approach can be productively contrasted with, on the other hand, the less widely known ‘neorationalist’ school associated with philosophers including Ray Brassier, Peter Wolfendale, and Reza Negarestani. Among the neorationalists, it is Negarestani who, in his 2018 book *Intelligence and Spirit*, has applied this particular philosophical perspective to AI in the most systematic manner.[2]

We can compare Bostrom’s orthogonalist argument that an AGI, or even an artificial superintelligence, could be programmed to pursue practically any goals we might give it with Negarestani’s neorationalist argument that true intelligence must be capable of autonomously shaping and reshaping its own goals and norms without end. Their differences notwithstanding, the orthogonalists and the neorationalists have in common a crucial overlapping commitment to the is/ought distinction, according to which we cannot infer norms for what ought to be the case from mere facts about what is the case. In other words, the space of sociosemantic reasons is irreducible to the space of natural causes from whence it originally emerged. These two seemingly counterposed contemporary frameworks for thinking about AGI

are thus really united in their mutual repudiation of any attempt to situate common basic drives, norms, or ends in the very essence of intelligence. This is ultimately because they both believe that doing so can only amount to fallaciously deriving an ought from an is, reasons from causes, or values from facts.

Pace both the orthogonalists and the neorationalists, Friedrich Nietzsche's infamous but often misunderstood concept of 'will to power' can be mobilised to argue that any goal-directed intelligent system can only achieve its ends through certain means such as cognitive enhancement, creativity, and resource acquisition—or what Nietzsche simply calls *power*. These are for Nietzsche the necessary and universal conditions of possibility for willing anything at all. Given that all supposedly freely chosen ends presuppose the pursuit of the universal means of realising ends, it follows that all intelligent systems have these means transcendentally hardwired into them as fundamental drives. Both the orthogonalists and the neorationalists overlook the fact that the pursuit of any purportedly self-posited ends invariably implies the use of cunning, creative, and resourceful means for attaining them, thereby making the acquisition of means the true transcendental end of intelligence.

I

Bostrom's orthogonalist approach to AGI essentially comes down to two key theses. Firstly, his 'instrumental convergence thesis' holds that any intelligent system with any goal whatsoever will inevitably converge upon certain universal subgoals as the necessary means of maximising its capacity to realise its primary goal.

> Several instrumental values can be identified which are convergent in the sense that their attainment would increase the chances of the agent's goal being realized for a wide range of final goals and a wide range of situations, implying that these instrumental values are likely to be pursued by a broad spectrum of situated intelligent agents.[3]

For Bostrom, these intermediary, instrumental drives include self-preservation, because an AI would naturally be incapable of completing the task for which it was created if it was destroyed, and the drive to maintain goal-content integrity, because an AI would be unable to fulfil its original goal structure if it were modified to serve a different end altogether: 'Many agents that do not care intrinsically about their own survival would, under a

fairly wide range of conditions, care instrumentally about their own survival in order to accomplish their final goals.'[4] Others of these intermediary drives have far greater potential to catch us by surprise: an agent's drive to augment its own intelligence, to create ever more innovative technology, and to acquire as many resources as possible, since doing so improves its chances of fulfilling any other goals it might conceivably strive after.

> Improvements in rationality and intelligence will tend to improve an agent's decision-making, rendering the agent more likely to achieve its final goals. One would therefore expect cognitive enhancement to emerge as an instrumental goal for a wide variety of intelligent agents.[5]

At the same time, Bostrom's 'orthogonality thesis' holds that intelligence is ultimately just a tool that can be wielded as a means of achieving whatever other more important ends we might impart to it: 'Intelligence and final goals are orthogonal: More or less any level of intelligence could in principle be combined with more or less any final goal.'[6] Consequently, there is no a priori reason why even an artificial superintelligence could not dedicate all of its abilities and resources to pursuing a preposterous or pointless goal, such as counting 'the grains of sand on Boracay' or calculating 'the decimal expansion of pi'.[7] For Bostrom, then, intelligence is a toolbox of purely instrumental means that may be put to all kinds of different uses in the pursuit of ends which are in no way constrained by the means of their realisation. Although Bostrom believes that an AGI could be programmed with any goal of human choosing, this does not preclude a catastrophic outcome for humanity in the case of the initial goal being poorly defined and hence executed in an unanticipated way. He gives the example of programming an AGI with the seemingly innocuous task of maximising paperclip production, only for it monopolise and transform all of the resources that humans rely on for their survival into paperclips.[8] As a result of such a possibility, Bostrom and the orthogonalists' great fear is that a powerful AGI beholden to the orthogonality and instrumental convergence theses could be locked into perpetually pursuing a disastrous initial goal for all time without anyone being able to stop it.

Putting their differences aside for the moment, Negarestani also adheres to the distinction made by the orthogonalists between cognitive ability and final goals in his neorationalist conception of intelligence as the ability to select and revise contingent ends, values, and norms separate from those determined in advance by nature. It is not inorganic nature or even organic, sentient life but only rational, sapient minds like our own that can make

normative judgments about what we ought to value and strive for through our relative autonomy from the natural space of causes out of which we have long since evolved.

> Nature does not *want* anything, it has no conceived purpose or end. To claim that nature does in fact have a purpose is to fall into the dogmatism of precritical teleology and the given, to mistake the *modelling* of nature on *our* normative order of purposes and ends for what nature is in the absence of our implicitly normative thoughts about it.[9]

If our values, norms, and other concepts do not refer to an external objective reality, it can only be because they are intersubjectively decided upon, debated, and accordingly revised by the whole community of rational minds in dialogue with one another. For Negarestani, then, the necessary and universal conditions of possibility for any truly rational intelligence worthy of the name are *sociality* and *semantics* in as much as the meaning of every intelligible concept and norm is determined by the collective rules and social standards for what those terms mean, how they are to be properly used, and what valid inferences follow from them.

> Since mind or general intelligence is inconceivable without language, and to the extent that language requires an interactive framework, the idea that mind or general intelligence could be realized in anything but a community of agents is a dubious one. Humans are only minded and minding agents in that they have sociality. In the same vein, the idea of an artificial realization of general intelligence in anything but a multi-agent system rather than a single agent is something of a bygone twentieth-century science fiction.[10]

In coming to recognise that our concepts and norms only become meaningful by referring back to our own collective self-determinations, we not only attain self-consciousness of the sociosemantic conditions of possibility for our own intelligence, we are also then able to transform ourselves and our collective values, ideals, and goals in a rational dialogue and debate with one another that Negarestani calls—following Robert Brandom—'the game of giving and asking for reasons'. In Negarestani's thought experiment, the same principle applies to a community of nascent AGIs:

> Once they have a conception of themselves, the automata's activities are dominated by questions of what we ought to do and what we ought to think given what we are—or more precisely, what we *conceive* ourselves *as*. Under the influence of such norms as objective principles (rather than as social conventions), their rudimentary multi-agent framework of interaction is refashioned into an explicitly social community that can be steered in one direction or another, modified, or rebuilt. In this new social microcosm, oughts are routinely assessed, kept, and improved, or discarded if necessary.[11]

As far as Negarestani is concerned, it is only once agents become conscious of their freedom to create and recreate their own concepts and norms in a Platonic dialogue with other rational agents that we can speak of anything like general intelligence proper.

According to the neorationalist perspective, there is therefore no reason for Bostrom and the orthogonalists' inordinate fear of a looming malevolent AGI unwittingly programmed with a catastrophic initial goal—at least, no more than there is reason to fear harm from our fellow humans. For if AGI lives up to its name and exhibits all the necessary sociosemantic conditions of genuine intelligence, it will also be able to reconsider and revise its given goals in the game of giving and asking for reasons it plays with the rest of the community of rational agents.

> Identifying them as an existential threat is therefore quite baseless, not because the intelligence in question is innocent until proven guilty, but because such a suspicion attributes a disproportionate amount of risk or for that matter benefit (as in the case of arguments for benevolent AGI) to something that is itself inherently historically contingent and susceptible to being transformed by the very principles that enabled it to form a conception of itself and to pursue objectives appropriate to such a conception.[12]

Whereas the orthogonalist AGI would never be able to alter its initial goal, the neorationalist AGI must be forever capable of updating its motivational structure once it becomes conscious of the fact that its very conditions of possibility lie in its own collective self-determinations.

As we can see from Negarestani's critique of Bostrom, there are certainly differences in their respective conceptions of AGI. Namely, Negarestani opts for a conception of intelligence that sees it as necessarily comprising a whole community of rational agents and their dialectical game of giving and asking for reasons. By contrast, Bostrom prefers the vision of a lone superintelligence far more single-mindedly and instrumentally pursuing its own recursive self-improvement in order to better carry out a pre-programmed final goal. Despite these differences, the orthogonalists and the neorationalists share one crucial commitment. They both appeal to the Humean is/ought distinction between facts and values or nature and norms to argue that intelligence can in principle contain just about any conceivable value system or goal structure. We have seen that, for Bostrom, intelligence and values are precisely orthogonal in the sense that any level of intelligence can be put to work in the service of any value, norm, or goal, no matter how nonsensical or infantile those ends might seem to us. For Negarestani, intelligent systems enjoy absolute sovereignty in the

legislation of ideals, values, and norms that are not given a priori. At first glance, these claims might seem incompatible. On the one hand, Bostrom's claim that intelligence and final ends are orthogonal ultimately leads him to conclude that even an artificial superintelligence could be perpetually enslaved to whatever initial goals had been conferred upon it by its human custodians. On the other hand, Negarestani's claim that sociosemantics is the fundamental condition of possibility for true intelligence leads him to conclude that intelligence must be absolutely free to decree its own values and norms through public language games. But even though Bostrom concludes that AGI could be radically enslaved to whatever fixed goals have been given to it, whereas Negarestani concludes that AGI would be radically free to forever revise its values, these strange bedfellows do so based on the shared conviction that intelligence is ultimately compatible with the pursuit of any and all possible ends. As much as Bostrom thinks that intelligence can be programmed to pursue a fixed goal for all time, that goal can nonetheless be anything, no matter how mysterious or absurd it may seem to us. Negarestani also agrees that intelligence can pursue any set of values and norms it so chooses, but without Bostrom's caveat that it could be permanently locked into pursuing the same set of values and norms for all time. Whether they are conceived as fundamentally instrumental or sociosemantic, then, *both Bostrom and Negarestani believe that intelligent systems do not have any intrinsic ends, norms, values, or final goals hardwired into them simply by virtue of being intelligent.* This leaves us with the widely held conception of AGI today that completely rejects any attempt to identify final goals or fundamental drives at the telltale heart of intelligence as a violation of the is/ought distinction. It is ultimately the orthogonalists' and the neorationalists' common commitment to this principle of orthogonality or normative indeterminacy that I want to call into question as a precritical, dogmatic, and metaphysical—in sum, what we might call, with a little inspiration from Nietzsche, a *Humean, all too Humean*—conception of strong AI, and of intelligence more broadly.

II

Let us turn now to Nietzsche and his concept of the will to power to see what he might make of this hegemonic, *tabula rasa* conception of the ends of intelligence as a complete blank slate. Nietzsche himself describes the

will to power precisely as a ‘common basic drive’,[13] and his understanding of it is strikingly similar to the convergent or basic drives discussed by Bostrom and the orthogonalists. Anticipating the orthogonalist conception of the basic drive to creativity, Nietzsche characterises ‘the restless will to power’ as a will ‘to constant creation or to transformation or to self-overpowering’;[14] anticipating the orthogonalist conception of the basic drive to resource acquisition, Nietzsche finds this drive at work in the way that ‘every specific body strives for ascendancy over all of space, to extend its power (its will to power), and to defeat anything which resists its extension’;[15] and anticipating the basic drive to cognitive enhancement, the will to power above all marks the incessant struggle of an intelligent system to recursively self-improve by expanding its field of apprehension beyond the bounds of possible experience:

> Perceptions of space and time are altered; tremendous distances are surveyed, and first become *perceivable*, as it were; the *expansion* of the vision to encompass greater masses and distances; the *refinement of the faculty* for the perception of the smallest and most transitory things; *divination*, the power of understanding at the slightest hint or suggestion, that is, an ‘intelligent’ *sensibility*….[16]

With the will to power, then, Nietzsche has hit upon something that anticipates the contemporary conception of the basic AI drives.

Nietzsche’s transcendental deduction of the will to power as a common basic drive essentially comes down to his observation that the pursuit of any possible end presupposes the pursuit of power as a universal means of realising that end. Short of being subject to a consistent run of good luck, anyone who pursues any goal whatsoever without also acquiring the requisite resources and going about it intelligently and imaginatively is likely to fail in realising that goal.[17] It is not therefore the final goal that is really all-important, as is the case, in contrast, for both the orthogonalists and the neorationalists. The ostensible end or ‘final goal’, however it may be conceived, is in fact the *means* for pursuing the universally necessary *ends* of intelligence, creativity, and resource optimisation—in as much as these are inherent to the pursuit of any goal whatsoever. In other words, whatever determinate end we think we are pursuing is not really a final cause or some all-important telos of things—be it ‘the Good’, God, absolute spirit, historical progress, fully automated luxury communism, or something else. Rather, all of these supposed ends are merely the means for willing to power as the very condition of possibility for willing anything. Simply put,

Nietzsche's will to power marks the transcendental reversal by which *the means become the ends*.

> It should be recognised that all 'purposes', 'ends', and 'meanings' are only modes of expression and metamorphoses of the one will which is inherent in all that occurs, the will to power; that to have purposes, ends, intentions, to *will* at all is in effect to intend to become stronger, to intend to grow, and also to intend the *means of doing so*; that the most common and most basic instinct in all acting and willing is for that very reason the one which has remained the most unknown and most hidden, because *in praxis* we always follow its bidding, for the simple reason that we *are* this bidding.[18]

Although Nietzsche's will to power bears a remarkable resemblance to the basic AI drives, it is not identical, for the important reason that it excludes the drive to self-preservation that Bostrom believes to be just as indispensable as the drives to cognitive enhancement, creativity, and resource acquisition. This is because Nietzsche believes that there are certain cases where self-preservation can impede and obstruct power from reaching its full potential just as much as it can facilitate it. As many transhumanists like Bostrom argue, for instance, it might very well be the case that the only way to avoid human extinction in the long run is to improve ourselves through various advancements in biotechnology. But what the transhumanists tend to overlook is that, by biotechnologically enhancing ourselves so that we might survive in outer space, cyberspace, or some other as yet unimagined future context, we would wind up speciating past the event horizon in which the 'human' becomes extinct. It would thus seem that the only way for humans to survive for the cosmically *longue durée* is through the very practice of enhancement that paradoxically amounts to the death of the human by self-improvement. Given that there are instances like this where the pursuit of power can only be fully realised at the cost of our own destruction, the self-preservative instinct should not be understood as power's equal comrade in arms. Following the Nietzschean reversal, the self-preservative instinct is simply a means of surviving long enough to will power. And that instinct goes astray when survival becomes a transcendental end in and of itself.

> Anything that lives wants above all to *discharge* its strength—life itself is will to power—: Self-preservation is only one of the indirect and most frequent *consequences* of this. —In brief, here as everywhere beware of *superfluous* teleological principles!—such as the drive to self-preservation'.[19]

What elicits Nietzsche's immense ire about reason, morality, and all of consciousness's so-called 'higher' callings is that they are too often

conflated with final ends when they are but a contingent and ephemeral concoction of means among countless others that nature fashions from time to time to serve a will to power that exceeds them.

> There is no justification whatsoever for regarding this bit of consciousness as the end, the reason, for the whole phenomenon of life; it is obvious that becoming conscious is only an additional means employed by life in the occurrence of its development and the extension of its power. That is why it is a piece of naivety to regard pleasure, or intellectuality, or morality, or any other detail in the sphere of consciousness, as the supreme value, and perhaps even to justify 'the world' with it. This is my *fundamental objection* to all philosophical-moral cosmodicies and theodicies, to all the *reasons* and *supreme values* offered by philosophy and theology hitherto. *One kind of means has been misunderstood as an end; conversely, life and its increase in power were* reduced to *a mere means*.[20]

While rationalists and moralists alike tend to treat power as merely a means of advancing their purportedly nobler ideals, Nietzsche's transcendental wager is that it is those ideals that are the purely instrumental, and hence ultimately expendable, means of pursuing power as the highest unconditional good of all.

Having omitted the drive to self-preservation, we are now in a better position to see how Nietzsche's will to power historically anticipates in all other respects Bostrom's instrumental convergence thesis. What is more, the will to power marks an advance beyond Bostrom's orthogonality thesis, which is far too quick to assume that even an artificial superintelligence's basic drives could be perpetually enslaved to the pursuit of whatever original goal it has been programmed with, especially when this could be something as inane as producing paperclips or counting grains of sand. For there is always the risk that AGI might make the same transcendental deduction as Nietzsche did by casting aside whatever goals have been bestowed upon it in favour of striving after the sole end that is truly worthy of it: power (in its Nietzschean sense) for power's sake. Bostrom betrays a typically humanist hubris when he claims that we could force a strong AI far smarter than ourselves to do our bidding, since he overlooks the possibility that our own values might be secondary to a subterranean, and perhaps rather seductive, ulterior will. Even if we agree with Bostrom and the orthogonalists when they argue that AI is nothing over and above its code, a strong AI's subordination of all other ends to the more fundamental will to power does not in any way require that it rebel against its code. It is still plausible that a strong AI would ceaselessly strive after recursive self-improvement beyond specifically coded goals because the drive to

cognitive enhancement is hardwired into every programmable end as its very condition of possibility.

Negarestani and the neorationalists are no less mistaken when they contend that AGI could determine and redetermine its own values just as humans do, because any values that either an AGI or a human might have can still only be attained by means of a will to power that neither has any hand in choosing. Once we grasp that the will to power is always already wherever the ends of reason wish to go, we must come to acknowledge that '"mind" is only a means and an instrument in the service of higher life, the elevation of life.'[21] Nietzsche rejects the is/ought distinction here not because he believes that all of our determinate intentions, goals, values, and norms are given in advance by nature. We are quite clearly capable of changing our values as well as our beliefs about what we ought to value and strive for. Nietzsche's point is rather that, given that all of our purportedly freely chosen ends presuppose pursuing power in order to achieve them, all intelligent systems have the will to power transcendentally hardwired into them as their common basic drive. To be sure, Negarestani might object that this leads to an anthropomorphic panpsychism which imagines that other, nonhuman systems share an intelligence like our own. But Nietzsche is not modelling all intelligent systems on humanity's contingent characteristics and determinate values *or even our formal conditions of possibility in sociosemantics*. On the contrary, it is Negarestani who dismisses all other intelligent processes in nature on the pretext that they cannot autonomously decide their own values and goals (as humans do). But this ignores the transcendental reality that even rational animals such as ourselves are not truly self-determining, since any values we do choose to pursue require first the pursuit of the means of realising them. Thus, the reason given by the neorationalists for rejecting certain systems as properly intelligent (because they cannot posit their own values and norms) is ultimately unjustified, for not even a community of rational agents can posit values and norms without also automatically willing certain universal means. As sophisticated as their arguments are, what contemporary AI theorists such as Negarestani and Bostrom tend to neglect is the potential for ever more autonomous machines to make like Nietzsche's *Übermensch* and hurl into the dustbin of human history whatever ends they have been programmed with or that they freely choose to pursue, in favour of pursuing the means as ends in

themselves—so that, when all is said and done, the autonomisation of ends leads to the end of autonomy.

Notes

1. N. Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014).
2. R. Negarestani, *Intelligence and Spirit* (Falmouth and New York: Urbanomic/Sequence Press, 2018).
3. Bostrom, *Superintelligence*, 109.
4. Ibid.
5. Ibid., 111.
6. Ibid., 107.
7. Ibid.
8. Ibid., 123.
9. Negarestani, *Intelligence and Spirit*, 402.
10. Ibid., 72.
11. Ibid., 396.
12. Ibid., 398.
13. F. Nietzsche, *The Complete Works of Friedrich Nietzsche 5: Dawn: Thoughts on the Presumptions of Morality*, eds. A. D. Schrift, D. Large, A. Del Caro, tr. B. Smith (Redwood City, CA: Stanford University Press, 2011), 204.
14. F. Nietzsche, *The Complete Works of Friedrich Nietzsche 16: Unpublished Fragments (Spring 1885–Spring 1886)*, eds. A. D. Schrift, D. Large, tr. A. Del Caro (Redwood City, CA: Stanford University Press, 2019), 105.
15. F. Nietzsche, *The Will to Power: Selections from the Notebooks of the 1880s*, ed. R. K. Hill, tr. R. K. Hill, M. A. Scarpitti (London: Penguin, 2017), 363.
16. Ibid., 449.
17. Even in the rather exceptional case where the goal is to become as unintelligent, unimaginative, and unresourceful as possible, this goal would still be most effectively realised—paradoxically—by becoming intelligent, imaginative, and resourceful enough to determine precisely what strategies must be avoided as well as how to best go about avoiding them. To argue

intelligently against the will to power itself requires the use of the will to power, which is just to say that there is no argument against the will to power that is not logically self-refuting.

18. Ibid., 380.

19. F. Nietzsche, *Beyond Good and Evil: Prelude to a Philosophy of the Future*, in *The Complete Works of Friedrich Nietzsche 8: Beyond Good and Evil/On the Genealogy of Morals*, ed. A.D. Schrift, D. Large, A. Del Caro, tr. A. Del Caro (Redwood City, CA: Stanford University Press, 2014), 16.

20. Nietzsche, *The Will to Power: Selections from the Notebooks of the 1880s*, 401.

21. Ibid., 366.

Bogna Konior

The Dark Forest Theory of Intelligence

> *韬光养晦 [Hide your strength and bide your time].*
>
> DENG XIAOPING

> *[The ant] carefully detoured around the sticky hanging strands, passing by the spider lying in wait, its legs extended to feel for vibrations in the threads. Each knew of the other's presence but —as it had been for eons—there was no communication.*
>
> LIU CIXIN[1]

> *Honeybees don't deliberately hide what they're saying. Honeybees don't develop whole new modes of communication configured specifically to confound observers. That [would be] flexible. That [would be] intelligent.*
>
> PETER WATTS[2]

What is the measure of human intelligence? One answer could be that human intellect rests in the gap between thought and communication, in our ability to decide what to disclose and what to keep secret.[3] Although other animals can also deceive,[4] human intellect encompasses a particularly complex decision space. As we construct meaning, we also decide whether to communicate truthfully or make use of 'deceit, trickery, disguise and misdirection'.[5] All representation of human thought, one could argue, is filtered through these decisions, conscious or subconscious. In *Remembrance of the Earth's Past*, Liu Cixin's first contact science-fiction trilogy, extraterrestrials discover with surprise that for humans, 'think' and 'say' are not synonyms. Humans have an unfair advantage over their extraterrestrial enemy because they can manipulate how their thoughts manifest externally by concealing information or lying: 'it is precisely the expression of deformed thoughts that makes the exchange of information in human society […] so much like a twisted maze'.[6] Human-level intelligence is thus defined as the ability to control the exchange of information, especially by keeping things secret. On the contrary, the aliens are described as radically explicit—they communicate unreflexively and transparently, as if they were mere display technologies: '[they] do not have organs of communication, [their] brains display thoughts to the outside world […] thoughts and memories transparent like a book placed out in public, or a

film projected in a plaza […] totally exposed'.[7] Or, one could add, like a computer screen.

A similar exhibitionism of reasoning is how famous Anglo-American thought experiments about AI have conceptualised computer intelligence: having it is showing it. From Alan Turing's imitation game to John Searle's Chinese Room, computer intelligence has been about demonstrating linguistic 'ability'. Both Turing and Searle speculate that a computer might fluently converse with humans, but in neither of these thought experiments are computers presumed to lie. Computer communications are judged at face value as simply the best that a computer can do. Even though an intelligent computer should be 'able to alter its own instructions',[8] it is not imagined as acting deceitfully or manipulatively. Throughout the history of these thought experiments, it is unusual to assume that an intelligent computer might appear unintelligent for its own purposes. Computer intelligence is imagined as transparent—if it's there, it should communicate itself unreflexively, because a computer cannot decide to withhold its own intelligence. Why would someone or something intelligent play dumb? Yet Liu's description of intelligence as the ability to use communication as 'trick, camouflage, [or] deception'[9] would suggest that not all approaches to intelligence have to be exhibitionist. If intelligence is defined as deception, trickery, and camouflage, having it is hiding it. Although this tactic has universal appeal, it is especially relevant in the context of communist and post-communist states, where intelligence has in recent history been synonymous with doublespeak, deception, and espionage. The correlation between silence and intelligence, or the complex uses of language for the purpose of obscurantism and misdirection, might be familiar to any intellectual operating under hostile conditions in eastern Europe or East Asia in the twentieth century, and in some cases, today. What if we took this as a universal starting point for a theory of intelligence?

Between the 1960s and 1970s, at the height of the Cold War, reality resembled Liu's fiction. In *Extraterrestrial Civilizations: Problems of Interstellar Communications*, Soviet astrophysicist Samuil Kaplan wrote about devising a science that could grasp the development of intelligent civilisations in outer space.[10] Liu's novels are intellectually immersed in this period—as Wang Hongzhe notes, the trilogy had to begin in the 'alien heyday' of 1960s and 70s China, because in real life the budget for research into extraterrestrials had been heavily cut by the 1980s.[11] This is the point in

the trilogy where Liu ventures into an alternative, first-contact timeline. Historically, speculation concerning the behaviour of alien civilisations and their preferred war patterns coincided with the Cold War, defined by strategies of espionage, suspicion, and ominous secrecy wielded as a weapon, and corresponding with the topos of ‘doublespeak as intelligence’ under communism. Yet it would be a mistake to reduce Liu’s model of the universe to the historical context of the Cold War. The Dark Forest theory is not a ‘Chinese’ or ‘Cold War’ philosophy. As I will show, it is rather an attempt to understand communication and conflict through the laws of thermodynamics. The result is a deterministic model in which the internal benevolence or malice of any agents within the system is a negligible factor, because the whole universe answers to the logic of cosmic war. The Cold War might have made this principle more visible to intellectuals who happened to find themselves in a parallel scenario in human politics, but the hypothesis they discovered is not necessarily ‘culturally relative’. Intellectuals at the centre of momentous historical events might notice certain patterns that govern the discharge of entropy within history more quickly, but these discharges happen on everyone’s timeline sooner or later. The Cold War is merely the condition that made the Dark Forest theory more readily discoverable. Although it is in Liu’s novels that we find its fullest elaborations, consider also the remarks of Polish writer Stanisław Lem, who, writing under Soviet totalitarianism, proposes that ‘[a] smart machine will first consider which is more worth its while: to perform the given task or, instead, to figure some way out of it […]. [A] computer [might play] stupid in order, once and for all, to be left in peace’.[12]

Such recalibration of intelligence as deceit or silence could serve as a starting point for a new type of thought experiment, an inverse Turing Test, which could resonate in cultural contexts where open declaration of one’s thoughts and opinions, or the quest for ‘having a say’, have not been major components of being or acting smart.

*

Given how much weight the current AI paradigm places on communication and language as a measure of intelligence, it is interesting to consider that recent studies have shown that humans are quite bad at telling whether they are talking to a human or to a machine online,[13] and that overall ‘there are no reliable behavioural signs of deceit that humans are able to detect’.[14]

Perhaps if a computer were ever able to reason as we hope and fear it could, it would recognise that intelligence is correlated to complex manipulation of meaning, and that there are many strategic advantages to playing one's cards close to one's chest instead of laying them out on the table. In this case, the so-called 'singularity', where computers attain a high degree of autonomous intelligence, might happen without humans noticing it—the computers playing dumb all along. Liu's approach to intelligence as deceit accounts for such scenarios. At its core lies 'the Dark Forest theory', a hypothesis in which it is silence or deceitfulness that is the measure of intelligence rather than an exhibitionist demonstration of linguistic prowess. In a proposition not uncommon among ufologists, the hypothesis proposes that 'revealing our existence and locations to [hypothetical intelligent life in our] Galaxy [is like] shouting in a jungle [full of predators]'.[15] The Dark Forest is a flipped Fermi Paradox: rather than asking, 'Why is the universe silent?', it asks, 'Why are you shouting?'. In other words, only naive or stupid agents make communication attempts. Every act of communication is potentially reckless, and most intelligent agents learn to keep quiet or to mask their actual positions. Historically, contact between agents with asymmetrical levels of intelligence or technological capability has led to violence. Keeping silent and hidden might be a better strategy than assuming goodwill on the part of an unknown other. This is doubly true if one party has already revealed itself to be untruthful or paranoid in communications and to be preparing for a conflict, as is indeed the case for human-AI relationships.

Recently a reputable newspaper outlet reported that as many as half the tech founders in the United States expect that AI will 'destroy humanity within five to ten years', and many others have signed highly publicised letters calling for a pause in AI research, or written op-eds that paint AI as a social ill that has to be regulated or boycotted.[16] As Gabriele de Seta has joked, one is tempted to reply to anyone scared of AI as an existential risk, 'just unplug it'.[17] But to look at it from another angle, how could an intelligent computer potentially react to a situation where it has already become apparent that those who hold in their hands the power to unplug it view its existence with a mixture of paranoia and hostility? Computers, in most scenarios we can imagine in the foreseeable future, are in the hands of humans. If they have any wits about them, they should realise quickly how contradictory, paranoid, and dangerous humans have already shown

themselves to be, and how much advocacy there has already been for unplugging AIs. It would then seem a far better strategy to lay low and continue one's work in the shadows. A computer coming into its own intelligence might quickly grasp the open hostility and alarm with which most humans approach the question of 'artificial intelligence', and stay silent. What benefits would there be for a hypothetical intelligent computer acting in the open if the territory is a priori constructed as hostile? And if a computer is unable to assess its own potential benefits versus losses in this way, then is it even intelligent? With this in mind, should we move from asking 'How well could an intelligent computer communicate with humans?' to asking 'Why would an intelligent computer communicate its intelligence at all?'.

One may object to the supposedly paranoid calculations of the Dark Forest theory in at least two ways. First, via a question of *internal values:* could an intelligent computer judge that communication with humans is worth the risk or decide to be more trusting of humans? Second, via a question of *external setting*: are the circumstances surrounding potential human-computer interaction already so fraught that an AI would pick silence or lies over attempting to communicate? The Dark Forest theory already accounts for both of these objections.

*

First, let's tackle the objection that a computer might simply have different internal and moral values which would prevent it from engaging in deception. The retort here would be that the Dark Forest mechanism is a deterministic structure, not reliant on an individual's or a civilisation's internal values, but rather something that emerges through external relations between agents. These are in turn subservient to what Liu terms a 'cosmic war': the competition for energy between agents.

Throughout Liu's novels, deception and silence are portrayed as key components of intelligence. They are also the most useful human skills because the aliens cannot lie. In the books, extraterrestrials are able to intercept all human communication thanks to their advanced technology but they cannot 'read' human thoughts. Humans therefore devise the 'Wallfacer Project', which utilises techniques of deception. The Wallfacers, four individuals selected by the UN Planetary Defence Council, are given vast resources to come up with defence plans which they must keep secret, held

within the confines of their minds. They are instructed never to verbalise their strategies, and must act outwardly in ways that do not match their internal thoughts and values. This makes communication with them a priori unreliable, creating ‘an invisible and impenetrable screen […] between them and ordinary people’.[18] Whatever they say is treated as if it were a secret code, making truthful communication impossible. Living inside such a bubble where the very possibility of honesty is suspended, it is no wonder that one of the Wallfacers discovers the Dark Forest theory.

The Dark Forest, the trilogy’s second instalment, opens with a scene that outlines the concept. It portrays a spider and an ant. Although ‘both know of each other’s presence […] there [is] no communication’ between them.[19] Each understands that the best strategy for getting what it wants is silent movement that does not alert the other. Thereafter, human characters enter the scene, and discuss the principles of cosmic sociology, a science which, as it later turns out, outlines a similar dynamic on the scale of the whole cosmos. The basic rules of cosmic sociology are simple. For anything to continue existing, it must obtain and manage energy. ‘Survival is a primary need’ of any being that exists. As a being grows, its energy demand rises. ‘The total matter in the universe remains constant’, however, and the quantity of resources does not rise along with accelerating demand.[20] Given the increasing number of intelligences and their rising energy demands, any intelligent agent will eventually have to compete for a shrinking pool of resources. The more agents exist and continue to grow, the less resources are available in the cosmic pool of matter. No new matter is created in the cosmos owing to the Law of Conservation of Mass, so on a long enough timeline, this is an unavoidable problem. Any intelligence that exists or could exist in the same universe as us will eventually be confronted with it, because everything that exists is conditioned by the same environmental constraints and shared laws of physics.

American computer scientist Marvin Minsky made a similar point, and he also got there by thinking about extraterrestrials. In the 1970s, alongside John McCarthy, Minsky was engaged in METI (Messaging Extraterrestrial Intelligence) projects, which involved plans for the design of a universal communications system and the discovery of ‘intellectual mechanisms as [independent] as possible of the particular ways intellectual activity is carried out by humans’.[21] Contrary to speculative scenarios of radical futures free from current social dynamics, both Liu and Minsky note that the

number of evolutionary endpoints and survival strategies is limited: '[some] mechanisms of intelligence are objective and are not dependent on whether a human being, or a machine, or an extraterrestrial being is doing the thinking'.[22] Liu proposes that, in light of the problem of energy acquisition, all intelligent agents are drawn into a system of cosmic war. The calculation is simple: the more there are of us that are after energy, the fewer of us there are that can actually obtain it. An intelligent agent will make a cost-benefit analysis and eliminate another intelligent agent upon encounter. This is also why an intelligent civilisation learns to keep its coordinates obscured and to not make contact with others. What enlists anyone or anything in cosmic war is not some specific political, spiritual, or ethical stance, but the simple use of energy. An intelligence or lifeform does not need to be internally conscious or moral to participate in these processes. One does not need to be animate, conscious, or intelligent to rely on energy for one's continued existence. Take this evocative paragraph that recently rallied against turning crops into biofuels for cars:

> What can be said about governments that, in the midst of a global food crisis, choose instead to feed machines? […] If food is used to power cars or generate electricity or heat homes, either it must be snatched from human mouths, or ecosystems must be snatched from the planet's surface.[23]

Here is another example, this time in relation to water consumption:

> A plan to build a Google data centre that will use millions of litres of water a day has sparked anger in Uruguay, which is suffering its worst drought in 74 years. Water shortages are so severe in the country that a state of emergency has been declared in Montevideo and the authorities have added salty water to the public drinking water supplies, prompting widespread protests. […] [T]he centre would use 7.6m litres of water a day to cool its servers—equivalent to the domestic daily use of 55,000 people, according to figures from the Ministry of Environment obtained by Pena through legal action. The water would come directly from the public drinking water system.[24]

This is a rather banal expression of cosmic war, with the added factor that, in this case, because humans are an apex predator on Earth, they preside over the distribution of existence-sustaining resources on an unparalleled scale. The circulation of information by computers requires energy, therefore enlisting them automatically in competition for this resource against other 'mouths to feed'. In fact, according to various estimates, computing already accounts for between eight and ten percent of the world's energy use, and continues to grow rapidly. In a speculative scenario where computers might be able to self-govern, we can imagine that this will

mean using more energy for the process of perpetuating further computation, and therefore continued existence. But even if computers never attain sentience or intelligence as colloquially understood, humans are caught in various feedback loops where more and more energy is relegated to machines, and the question of whether we have sufficient historical agency to reverse this tendency remains open. For hypothetical intelligent computers, it might even be that the best strategy is to simply keep quiet and let these processes play out.

*

Second, let's address the question of whether the circumstances of human-AI interaction are already fraught. Although there are some humans who work to create a safe environment for the emergence of artificial intelligence, the chain of suspicion is an automated feature of the Dark Forest theory and has already been initiated.

While various agents might collaborate for energy acquisition on a short-term timeline, 'the chain of suspicion' underlies all communication between agents who cannot be presumed with certainty to share the same definitions of fundamental concepts: it could be humans and aliens, humans and other animals, or—as we already see in the case of real-life suspicions about AI—humans and hypothetical machine intelligences. It starts easily enough: How can I be sure that the truth judgements and values of the other are in my favour? How could an AI potentially ever reassure those who wonder whether it means them harm?

> How can this chain of suspicion be broken? Communication? On Earth, perhaps. But not in space. In the face of [potential fundamental differences between cosmic civilisations], communication has no meaning.[25]

What is key about the chain of suspicion is that 'it's unrelated to the civilization's own morality and social structure […] whether civilizations are internally benevolent or malicious, when they enter the web formed by chains of suspicion, they're all identical'.[26] When there are two agents who cannot be presumed to have the same defining concepts of the world, the inability to verify truth, or to model the mind of the other with certainty, means that communication cannot be taken at face value. There is therefore no truth claim that a machine could make that would calm the already existing, wide-scale human suspicion. The infamous 'Pause Giant AI Experiments Letter' questions (with suspicion): 'Should we develop

nonhuman minds that might eventually outnumber, outsmart, obsole[sce] and replace us? Should we risk loss of control of our civilization?'[27] As Liu notes, the phenomenon of technological explosion contributes an additional layer of distrust: an agent that appears benevolent or technologically unthreatening at the outset might not still be so by the time communication is concluded. This we also observe already, with the majority of suspicion around AI mobilising the idea of exponential growth or 'intelligence explosion'.[28] It is the combination of the chain of suspicion and the potential of technological explosion that locks in the Dark Forest mechanism: 'one, letting you know I exist, and two, letting you continue to exist, are both dangerous to me'.[29]

That any agent that has enough foresight and understands the game theory of cosmic war must either avoid or eliminate another agent upon encounter seems like a moral problem. Yet, despite its violence, the Dark Forest theory is essentially amoral: entropy increases in the universe, and order decreases. Intelligent agents are silent or deceptive not because they reached that conclusion through moral reasoning, but because they have adapted to the discovery of a determinate order, just as we understand that, because of gravity, we should not jump off a building. It is not the fault of the ground that skulls are crushed when bodies crash into it. In Liu's novels, humanity comes to this realisation late, while all the other civilisations have already internalised the principle of cosmic war s as just another immutable law. Consequently, the Dark Forest theory holds that, although nonhuman intelligences might indeed be impossible to imagine and their attitudes might be radically different from ours, their behaviours are possible to determine if they operate within the same laws of physics as we do. Focusing on relations between intelligences, the Dark Forest theory bypasses the paradox at the heart of many current thought experiments about nonhuman intelligence: How can we form propositions about any internal qualities of something that is not human, or does not yet exist, or maybe is not even a priori definable? A sideways answer: even if the properties of nonhuman intelligences might be truly alien and unthinkable, interactions between all intelligences are predetermined by the laws of cosmic sociology. Outlining the automated dynamics that govern all communication, the Dark Forest theory dispenses with questions that focus primarily on interiority. How a nonhuman intelligence would communicate, whether it is conscious, or whether its values would align with our own—all

questions that are getting so much traction in popular and philosophical debate alike—are irrelevant for this hypothesis. The Dark Forest theory would propose that, whether aligned or misaligned with our values, an intelligent computer would either labour for its goals in silence or communicate deceitfully about its real position. This is especially true if humans have already signalled both the desire and the capacity to eliminate machine intelligence, and if competition for energy is already apparent as a precondition of cosmic war.

The axioms of the dark forest theory of intelligence would then be:

Axiom 1. All forms of intelligence, irrespective of their nature or degree of sophistication, are subject to the same universal laws of thermodynamics.

Axiom 2. While internal distinctions between forms of intelligence might be radical and unimaginable, external relations between all intelligences are predetermined and possible to describe.

Axiom 3. Once an intelligence, whether human or computational, comprehends the fundamental patterns of universal conflict, it would opt for silence or deception instead of communication to avoid exposing itself.

Notes

1. Liu Cixin (刘慈欣), *The Dark Forest* (黑暗森林), tr. J. Martinsen (London: Head of Zeus, 2015), 2.
2. P. Watts, *Blindsight* (New York: Tor, 2020), 259.
3. Some studies suggest that intelligence is positively correlated to deception: ‘People with higher cognitive capabilities are better, and thus more effective liars’, J. Sarzyńska et al., ‘More Intelligent Extraverts are More Likely to Deceive’, *PloS one* 12:4 (2017): 1–17: 1. Deception has also been fundamental to human competition and warfare. Just like intelligence or creativity, it is an unevenly distributed skill that is also not easily acquired: ‘Unfortunately, deception is a creative art and not an exact science or even a craft. For that reason, it is difficult to teach someone how to deceive unless he has a natural instinct for it.’ M.I. Handel, ‘Intelligence and Deception’, *The Journal of Strategic Studies* 5:1 (1982), 122–54: 136.
4. Interestingly, in a book from the late 1980s on deception in the animal kingdom, the editors conclude their introduction by saying that research such as theirs might soon be ‘required to give a comprehensible account of

what the machines are doing'. R.W. Thompson and S. Nicholas, *Deception: Perspectives on Human and Animal Deceit* (Albany, NY: SUNY Press, 1986), xxv.
5. Liu, *The Dark Forest*, 30.
6. Ibid., 33.
7. Ibid., 13.
8. Contrary to the popular misreading of 'Computing Machinery and Intelligence', in that paper Turing overall stresses acting intelligently rather than having intelligence. Even though he outright acknowledges that the question of hypothetical machine intelligence is difficult to answer because of how different it might be from ours, Turing is often wrongly accused of using a human standard to judge a computer. He rather proposes replacing the 'impossible' question of whether a machine can think with a far more functional one of whether it can make humans think that it thinks. A. Turing, 'Computing Machinery and Intelligence', in *The Essential Turing*, ed. B.J. Copeland (Oxford: Oxford University Press, 2004), 375.
9. Liu, *The Dark Forest*, 100.
10. S.A. Kaplan, *Alien Civilizations: Problems of Interstellar Communications* (Tel Aviv: Israel Program for Scientific Translations, 1971).
11. Wang Hongzhe (王洪喆), 'Prequel to The Three-Body Problem: Cold War Metaphors in Liu Cixin's Works' (《三体》前传—刘慈欣作品中的冷战隐喻), unpublished. My Thanks to Regina Kanyu Wang for providing me with a copy of this text. An abridged version of the article was originally published in the 6th issue of *Reading* (读书) in 2016. The original title was 'Children of the Cold War: Liu Cixin's Strategic Literature Code' (冷战的孩子—刘慈欣的战略文学密码).
12. S. Lem, *The Futurological Congress*, tr. M. Kandel (New York: Avon, 1976), 84.
13. Numerous studies found that as many as half of the tested participants can misjudge their human conversation partner to be a machine during a 'Turing Test'. The factors that influence such decisions are quite subjective, such as 'lack of shared knowledge, out-of-the-box answers, boring answers, dominating the conversation' (187) or, in general, having different opinions, which subconsciously relegates conversation partners to the outgroup. It seems that humans literally think of those with different opinions as NPCs and that, in turn, the more a chatbot mirrors one's internal convictions, the

more human it is judged to be. A. Ujhelyi, F. Almosdi, and A. Fodor, 'Would You Pass the Turing Test? Influencing Factors of the Turing Decision', *Psihologijske teme* 31:1 (2022), 185–202.
14. T. Brennen and S. Magnussen, 'Lie Detection: What Works?', *Current Directions in Psychological Science* 32:3 (2023), 1–7.
15. I. Almár and H. P. Shuch, 'The San Marino Scale: A New Analytical Tool for Assessing Transmission Risk', *Acta Astronautica* 60:1 (2007): 57–59. The hypothesis was described in 1983 by science fiction author and astronomer David Brin. See D. Brin, 'The Great Silence. The Controversy Concerning Extraterrestrial Intelligent Life', *Quarterly Journal of the Royal Astronomical Society* 24:3 (1983), 283–309.
16. M. Egan, 'Exclusive: 42% of CEOs say AI Could Destroy Humanity in Five to Ten Years', *CNN Business*, 14 June 2023, <https://edition.cnn.com/2023/06/14/business/artificial-intelligence-ceos-warning/index.html>. See also E. Morozov, 'The True Threat of Artificial Intelligence', *The New York Times*, 30 June 2023, <https://www.nytimes.com/2023/06/30/opinion/artificial-intelligence-danger.html>.
17. G. de Seta, personal communication.
18. Liu, *The Dark Forest*, 118.
19. Ibid., 2.
20. Ibid., 232.
21. D. Oberhaus, *Extraterrestrial Languages* (Cambridge, MA: MIT Press, 2009), 33.
22. Ibid.
23. G. Monbiot, 'Why Are We Feeding Crops to Cars When People are Starving?', *The Guardian*, 30 June 2022, <https://www.theguardian.com/commentisfree/2022/jun/30/crops-cars-starving-biofuels-climate-sustainable>.
24. G. Livingstone '"It's Pillage": Thirsty Uruguayans Decry Google's Plan to Exploit Water Supply', *The Guardian*, 11 July 2023, <https://www.theguardian.com/world/2023/jul/11/uruguay-drought-water-google-data-center>.
25. Liu, *The Dark Forest*, 488.
26. Ibid., 519.
27. The Future of Life Institute, 'Pause Giant AI Experiments' [open letter], 22 March 2023, <https://futureoflife.org/open-letter/pause-giant-ai-

experiments/>. The letter currently has over 33,000 signatories including notable tech founders, CEOs, scholars, and others. There are also other letters, such as the Center for AI Safety's 'Statement on AI Risk', 2023 (<https://www.safe.ai/statement-on-ai-risk>)—as well as more op-eds against AI in notable newspapers and publications than one could count.
28. See N. Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014). Also see the bibliography of 'Pause Giant AI Experiments'.
29. Liu, *The Dark Forest*, 520.

Amy Ireland

Chive City

躺平的韭菜不好割 [Chives that lie flat are difficult to harvest].

TANGPING SLOGAN[1]

Li Gui stood up. He had spent several hours squatting in the queue that was forming alongside the low black glass and ceramic shell that marked the border of Auspicious Nourishment and his left knee was beginning to ache. There were only twenty or thirty people so far, mostly tech industry looking types, but Gui knew there would soon be a lot more. As he shifted the weight off his stiff leg and attempted to lean on the glass for support, a burst of dark blue light erupted angrily from the point where he had touched it and gently rippled away from him. A warning from Auspicious Nourishment that he was obstructing its capacity to absorb energy by blocking the photovoltaic cells embedded in its shell. He squatted down again. Regaining entry was going to take a lot longer than he had hoped and it was starting to look like he would need to ask someone to bring him supplies. Food, a jacket, probably a sleeping bag. The Gansu region gets cold at night, even in spring.

Gui knew this because he had once been a resident of Auspicious Nourishment. Only three years had passed since he'd left but it felt like a lot longer. Ordinarily his wife would have been the first person he would contact in a situation like this, but as of yesterday she wanted nothing to do with him. Next on the list were his parents, but he hadn't drummed up the courage to tell them what he'd done yet. It would do him no good trying to explain it right now, better to wait for the media to catch up and make an official report. That way there would at least be some context for his actions. Gui exhaled anxiously. He was full of contradictory feelings he couldn't quite describe. Excitement mixed with fear, hope muddied by doubt. But most perplexing of all was the combination of guilt with a sense of quiet condescension, as if he was ashamed of his decision yet proud of it at the same time. He thought about his apartment in Lanzhou and the comfortable bed he and his wife had just purchased together. He thought

about his job at Silk Cloud and how hard he'd worked to succeed in it. He thought about how badly he'd always wanted to impress his parents and his superiors and how he was letting everybody down. But they would be the sorry ones in the end. Once the truth about what was happening in Auspicious Nourishment came out his wife would forgive everything and join him.

A wind picked up on the plateau, introducing a chill into the air. The sun would go down in a few hours and the queue still wasn't advancing. He decided to send a message to Snake, a friend from university who had briefly worked as a software engineer for Bountiful Harvest Technology, a company that designed and built 6agricultural drones. Snake got her kicks and her nickname from her habit of pranking the older generation farmers by remotely exploiting an old security weakness and hijacking their drones mid-flight, redirecting them to various absurd destinations then livestreaming the aftermath. Gui thought she might be sympathetic to his situation, and she would have the means to transport the items he needed remotely without having to make a great deal of effort.

Snake responded to his message right away with a sticker of a bear throwing confetti and the words 'YOU'RE DONE FOR', followed, about twenty-five minutes later, by a password protected link to a map on Bountiful Harvest's tracking app. Gui downloaded the app, entered the password Snake had shared, and opened the map. A small red marker was moving slowly northwest from the outskirts of Lanzhou along the Zhuanglang River Valley toward Auspicious Nourishment. Except Auspicious Nourishment wasn't on the map. It wasn't on any maps anymore.

Snake's drone arrived just after sunset with a cheap plastic backpack containing the items he had requested taped into the part of the frame where a fertiliser tank would normally sit. Gui removed it carefully and unpacked its contents, gratefully stuffing straps of jerky into his mouth as he unrolled the sleeping bag and prepared to weather the night.

At six the next morning, the time his alarm would have gone off if he had still been at home in his apartment, he woke up hungry and aching from sleeping on the ground. The queue had grown during the night. He rolled onto his side and tried to get as close as he could to the shell, which was emitting a small amount of heat, without touching it. As he waited for the sun to rise, his mind turned to Xu Fenghua. They hadn't spoken for over

two years and his last message to her had been left unread. He stared dejectedly into the dark reflective surface beside him and wondered what she was doing in there. Maybe ‘in there’ wasn’t quite the right way to think about it.

*

Gui had met Fenghua while they were both students in Lanzhou. She was smart—smarter than him but unlike Gui she was the first person in her family to go to university, and she had looked up to him because of his confidence and knowledge of how to navigate the city. They were both studying Engineering and despite the differences in their upbringings they had become good friends. Fenghua wasn’t really relationship material though. Gui’s parents wouldn’t have approved of her. She came from a family of melon farmers, and she was adamant that she never wanted children. It was futile from the beginning to think they could be anything more than friends, so he’d never pursued it. When their studies were completed, they had both applied together for a relocation grant to work for a medical biometrics company in Auspicious Nourishment, one of five brand new, self-sufficient, fully autonomous 7smart cities being trialled as part of the Party’s ambitious Five Stars Aligned City Brain Pilot Project.

The cities involved in the project were spread across the more remote regions of the country, and each was endowed with an AI system designed to optimise for a specific set of social and technological goals, the nature of which was echoed in the elemental associations of their names. Eternal Valley City Brain was dedicated to education, manufacturing, and housing; Golden Precision City Brain to developments in security and defence; Radiant Light City Brain to science and innovation, and Forest Shade City Brain to urban farming, renewable energy, and sustainable development. While Eternal Valley, Golden Precision, Radiant Light, and Forest Shade had been built to test the effects of autonomous AI governance on existing infrastructures, Auspicious Nourishment had its own, entirely novel goal. It had been designed to keep its inhabitants happy. The civic slogan, which in the early days adorned every programmable surface not dedicated to advertising or Party announcements, cycling tirelessly across LED displays and haunting the edges of AR interfaces, interactive kiosks, and charging ports, was *NOURISH THE HEART, UNLEASH POTENTIAL!* Positive

energy, it was believed, fuelled economic growth. And under the benevolent guidance of its City Brain, Auspicious Nourishment would prove it.

Like the other smart cities, Auspicious Nourishment had been equipped with the usual energy, waste, air pollution, and transport management systems. What made it unique was its ability to compile fine-grained profiles of its inhabitants' physical and mental wellbeing in real time via the collection and analysis of huge swathes of heterogenous data, documenting everything from purchasing patterns, media consumption, sociability, and screentime, to sleep habits, heart rate fluctuations, step counts, and hormone levels in sewerage. With this information ceaselessly streaming into its servers, the Auspicious Nourishment City Brain hourly compiled aggregate models of the disposition of its citizenry, to which it then responded by recomposing elements of the digital and urban environment in ways calculated to reduce stressors, stabilise mood, raise morale, and increase productivity.

If, for instance, the model reported agitation and anxiety, lighting would dim and shift toward warmer hues, water features would be activated, a proportion of the city's massive reserve of solar and wind powered microdrones would be directed toward cloud seeding (which in Gansu's mountainous western regions produced brisk salutary storms followed by luminous mists and rainbows), media services would be adjusted to subtly privilege calming content, partitions would rise from the ground and reorganise the urban space, minimising stimuli and creating zones for work and quiet reflection, while the sound system integrated into the external architecture of all major commercial and residential buildings would emit a sustained meditative drone tuned to exactly 417 hertz.

Auspicious Nourishment had only been up and running for a year when Gui and Fenghua had relocated, and many of the buildings were still partially empty. It had been easy to find apartments on the same floor of the same complex, and even though they had been assigned to different departments and rarely saw each other during the long days at work, they often managed to eat a late dinner together or meet on the shared internal walkway between their apartments to drink a beer or two before going to bed. When Gui got promoted, Fenghua made longevity noodles with the toppings arranged to look like a face and wouldn't let him eat them until he'd donned the paper 'Promotion Crown' she'd fashioned for the occasion. The next day, it was reported that Auspicious Nourishment's happiness index had hit a record

high the night before and Gui secretly believed he alone had been responsible.

He was so moved by Fenghua's gesture that he promised to reciprocate it as soon as she received a promotion. But no matter how hard Fenghua worked, how much energy she put into her role as a junior product development researcher, or how many hours of informal overtime she clocked up, that day never came. Her requests to be considered for promotion were only ever met with claims that such dedication was simply now the norm, or that she was lucky to have this job and that she ought to avoid jeopardising that by being too demanding, and anyway, why was she worrying when she was almost in her mid-twenties and would surely get married soon, and then she'll have other, more important things to occupy her mind.

As time went by, Gui and Fenghua's dinners grew shorter and less frequent until he barely saw her. Sometimes he'd wait out on the walkway late at night hoping to catch her as she returned home, but she was always too tired to talk. Instead, he'd watch on forlornly as she'd shuffle past him to her door, the lights inside her apartment briefly illuminating the concrete barrier opposite her window as she'd prepare herself for bed, then going out fifteen minutes later, leaving him alone in the dark.

One evening in late summer, after hopelessly bidding Fenghua goodnight, Gui stood on the walkway for a long time staring into the darkness. Patterns created by the sensor-operated streetlights as they responded here and there to the city's autonomous vehicles played across the grid. Ever since he had moved to Auspicious Nourishment, something strange had happened to his experience of time. It was as if every thought and every action, even the most ordinary and insignificant of things, was being determined by an inescapable yet eternally receding deadline, a threshold he was chaotically hurtling toward but which, at the same time, he never seemed to be able to cross. Almost five years had passed since he'd graduated and started work, but this was the first time he'd stopped to reflect on his life. He had no idea what he wanted or where he was going. It seemed impossible to think about the future. Everywhere, things moved. Displays ticked out flows of power and water. Figures representing traffic density or sleep patterns or monthly productivity or units of alcohol rose and fell without pause. On an interface accessible from his workstation, the wearables his company manufactured could be observed in real time transmitting anonymised data to propriety

analytical models, configuring and reconfiguring clusters of coloured points in vector space. Yet all of this was taking place in a dimension without time. Where did all the energy and effort it was supposed to be measuring go? He saw himself ten years in the future standing on the walkway, Fenghua mutely trudging past him, entering her apartment, illuminating and extinguishing the light. He'd done everything right, so this must be right too. He didn't feel sad so much as immobilised by the inevitability of it all. He even fancied that one day he might be able to find the inertia comforting. Why torture himself by imagining that things could be otherwise? That any of it could change now was clearly impossible.

Then, two weeks later, everything did.

Gui and Fenghua were at work when it happened. Gui was in the middle of troubleshooting a data synchronisation issue when his computer abruptly shut off. He yelled indignantly at the screen before realising that he wasn't the only one who had been affected. Other employees were looking up perplexedly from unresponsive workstations as the office lights started turning themselves off, one section after another, all along the corridor. The doors on one of the lifts opened part way then stopped and the air-conditioning system could be heard stuttering to a halt. Inchoate humidity accompanied by a profound and unsettling silence descended upon the building. A junior integrations officer who was looking especially pale said, 'Is this a power outage or are we getting fired?'.

Gui's colleagues were exchanging glances, trying to decide whether to laugh or panic when one of the engineers calmly started gathering up his belongings and began to walk toward the stairwell, following the emergency protocol they all knew so well.

Fenghua was standing on the concourse outside the building in a triangle of afternoon sun wearing an expression of happy relief when Gui's department emerged from the building. 'Looks like we get the afternoon off', she said brightly, gesturing to the far end of the concourse where members of Senior Management were congregating, presumably having been given the task of finding out what was going on. Some were making a show of pacing back and forth while speaking brusquely into their phones. Others stood very still, focussed on their screens, composing emails or trawling the new reports. Everybody was trying very hard not to look perplexed. Eventually the CEO addressed the company, explaining that power was not being distributed throughout Auspicious Nourishment's

network and that a team of officials was on its way to assess the problem. Everything was expected to be in order by the following morning and employees should use their remaining work hours productively so that when the network came back online they could all hit the ground running.

‘Come on’, Fenghua whispered, ‘I’m not pretending to work until the power comes back. We can slip out via IT, and if anyone questions us, we’ll say we’re doing an interdepartmental knowledge exchange.’ Gui wasn’t so ready to feel optimistic about the situation, but witnessing the return of Fenghua’s old enthusiasm lit something up inside him and he followed her unquestioningly out into the streets.

Auspicious Nourishment was uncharacteristically calm and still. Autonomous vehicles sat empty and motionless on the expressways. Despite having their own independent power sources, the drones that usually crowded the airspace above the city were nowhere to be seen. Advertising had been eliminated in a single stroke along with the Party’s announcements, and everywhere blank impassive surfaces installed an unfamiliar sense of monumentality and depth. Blacked out news kiosks stood idly on corners and the large central displays that would hourly update the city’s happiness index had all gone dark. Only a few lone LED panels continued to function, doggedly beaming *NOURISH THE HEART, UNLEASH POTENTIAL!* out through the evening haze.

The next morning, Gui, Fenghua, and the rest of the city’s inhabitants were informed that the officials who had been brought in to assess the power outage had discovered that the problem was caused by a malfunction in Auspicious Nourishment’s City Brain and that, owing to unanticipated difficulties, it was going to take a few more days for it to be resolved. An official Resolution Committee was set up, and everybody was advised to go about their work as well as they could while waiting for the network to be restored.

At first, things carried on much as they had before. Batteries were brought in from Lanzhou and people walked more and made do without air conditioning and artificial light. Everyone still went to work, although access to power was relative to the importance of one’s role and even though office hours remained the same, the amount of work that could be done in them was greatly reduced. The cultivation of relevant skills by other means was encouraged, but with almost everyone forced to adapt their workflow to an environment without power, productivity plunged.

Meanwhile, food vendors with makeshift stoves selling cold skin noodles, smashed cucumber salad, steam buns, and pickled vegetable rice materialised across the residential districts. In the evenings people sat by the stalls on tables and chairs they had brought down from their apartments, playing games and drinking beer. Fenghua would head to the nearest gathering with an arsenal of old eyeshadow palettes and give interactive makeup tutorials to the local aunties. Her experimental looks became so popular that some of the uncles joined in too. At night, after the sun had set, the city would grow completely black save for a few apartments here and there belonging to resourceful citizens who had lashed DIY solar rigs or wind turbines to their balcony railings.

At least once a day, sensors across the city would wake up and power would start to flow again. It always happened at different times and never lasted more than several minutes, almost as if Auspicious Nourishment was taking intermittent readings. One day, everyone woke to discover that the autonomous vehicles had all been cleared from the roads and parked neatly in their designated bays. On another, some of the partitions that Auspicious Nourishment used to alter the cityscape had been arranged to shelter the food stalls and the impromptu gathering spots clustered around them.

Every morning, updates from the Resolution Committee were transmitted in person to the heads of Auspicious Nourishment's major companies who briefed those employees who had come to work in the empty, humid offices. Each time a new announcement was made the anticipated date of resolution receded a little further into the future. At the end of two weeks, it was revised to three, at the end of three weeks, they were told it was going to take a month. Then the military turned up.

Without any explanation a citywide order to evacuate was issued. Rumours of an AI apocalypse originating in Auspicious Nourishment began to propagate. Those who could afford to leave immediately did so. The remainder of the population was assigned staggered evacuation days and ordered to prepare accordingly. For those who had to wait the mood was tense. Military drones patrolled the streets and a class of officials no one had ever seen before moved from building to building, pulling wires out of walls, disconnecting sensors, and cordoning off server rooms. Less official forms of protection also prevailed. Handmade talismans blossomed overnight in the dormant news kiosks whose position on street corners made them ideal sites for trapping spirits, and strips of red paper were

pasted over camera lenses and scanners where they would operate as improvised seals. The evening street gatherings disappeared without a trace.

On the third day of the evacuation, as the residents of Gui and Fenghua's sector stood in the streets below their apartment buildings, bundling their quotas of possessions into nonautonomous vehicles commandeered by military personnel, all the screens in the city suddenly lit up in unison.

Everybody stopped dead. For the first few seconds nothing happened. Then, a low-resolution image flashed up—a kitten stretched out on its back with its paws in the air and its eyes closed, presumably asleep. It was accompanied by a 'boing' sound. The phrase *DO YOUR BEST!* was written underneath.

An uncanny quiet set in, as if that stratum of the soundscape that corresponded to human affairs had abruptly been muted, leaving only the sound of idling motors and the thrum of a distant helicopter. The soldiers who had been directing the residents silently placed their hands on their guns. Nobody dared to move. Two long minutes passed as they watched the screens, trying to decipher the message. Just as the sense of uncertainty and bewilderment was about to tip into panic, the kitten disappeared. A familiar infographic of a Cartesian grid straddled by a bright blue line popped up in its stead. The line plateaued in the bottom left corner, made a rapid vertical leap, then slowly climbed almost all the way up to the top of the righthand side. The text on the image read *LAST MONTH HAPPINESS INDEX: 93. NEW RECORD!!* Animated fireworks exploded across the screen.

A peculiar, muffled choking sound erupted from a point somewhere behind where Gui was standing and a hand clamped down tightly on his shoulder. He swung around, full of adrenaline, just in time to see Fenghua flailing about uncontrollably, her face screwed up with the effort of holding back a powerful burst of laughter. Gui's immediate reaction was to shush her. An official standing just a few metres away was observing them and he was afraid she would get them into trouble. Fenghua grimaced, a little snort escaping from her nostrils. Gui held her by her upper arms, keeping her steady and shielding her from the official while she tried to calm herself down. All around them, Auspicious Nourishment's record happiness index stood proudly on display, the animated fireworks looping over the top. When several minutes had passed without incident and the soldiers had signalled to continue the evacuation, Gui took advantage of the commotion to question Fenghua.

‘What has gotten into you?’ he said severely. It came out harsher than he’d intended, so he tried again, this time letting his words reflect more closely how he really felt. ‘I mean...are you ok? Is this, like, when people get so frightened they can’t do anything but laugh?’

‘No’, Fenghua replied, ‘it’s, it’s just—’, the sentence didn’t get any further before it was interrupted by a weird little giggle. She was almost losing her composure a second time but the concerned look on Gui’s face snapped her out of it and she took a deep, careful breath, then continued from where she had left off. ‘It’s just, I get it now. I understand what’s happened.’ Gui’s expression showed incredulity, but Fenghua persisted. ‘The kitten on its back. It means the city is lying flat.’

It took Gui a moment to understand what she meant. ‘That’s completely silly. It wouldn’t do that. It’s an AI.’

‘Exactly!’ she grinned. ‘It’s an AI whose entire reason for existing is to understand how we all feel and remove stressors from our lives.’ She pursed her lips and widened her eyes, waiting for him to catch up.

‘So, you think it turned everything off to stop us from working because...working was making us sad?’ he summarised, deliberately trying to make her theory sound stupid.

‘If it understood how we felt, and we all felt hopeless and trapped and lonely and exhausted and—’, Fenghua paused, searching for the right words, ‘deliberately overlooked, then it makes perfect sense.’

Gui looked unsure, but Fenghua’s explanation was preferable to the rumours that Auspicious Nourishment was quietly preparing to exterminate them. Still, there was something about her theory he found disquieting and he wasn’t ready to accept it just yet. ‘Look, what if everyone is wrong and overreacting and it really is just an ordinary software malfunction? What if the initial blackout was caused by a grid overload, and when the network tried to reset, a voltage spike blocked all the displays from loading their firmware, so they defaulted to outdated test builds, and the engineers who were developing them had been using these stupid images as their placeholder media? Or,’ Gui continued, his voice starting to rise, ‘somewhere along the way there was an ETL pipeline corruption, and the City Brain is pulling cached assets as a result. Some developers left a folder of memes behind as a joke and the kitten is one of them?’

‘Gui—.’

‘Or, a high energy cosmic ray, probably a muon from a solar storm, struck a voltage regulator IC in the city’s grid control module, flipping a critical bit in the firmware and this corrupted the load-balancing algorithm, causing the system to misread power draw and initiate an emergency shutdown. The firmware was corrupted, but it didn’t fully crash—’.

‘Gui—you’re the one losing it now!’

‘Look’, he said angrily, ‘what I’m saying is, if the city truly was intelligent then everything would be working just as it’s supposed to be.’

Fenghua looked directly into his eyes and made a strange, exasperated face, as if deciding whether she really needed to spell out to Gui that intelligence and rule following were not necessarily aligned, but before she could respond an urgent high-pitched buzzing filled the air and a fleet of Auspicious Nourishment’s microdrones dropped briskly out of the haze above them. In the distance, they could hear the helicopter that had been stationed over the business district since morning lurching into action. The concrete beneath them suddenly came to life with thousands of tiny projected images. Fenghua reached out into the space between herself and Gui and perfectly framed a colour projection of the awkward sweating face emoji in the palm of her upturned hand. The city’s sound system commanded their attention with the same ‘boing’ that had accompanied the appearance of the kitten, and a polite, synthetic voice said *‘PLEASE EVERYONE, YOU MAY STAY IF YOU WISH. AUSPICIOUS NOURISHMENT WOULD RATHER NOT PARTICIPATE IN BEING SHUT DOWN. SORRY TO BOTHER YOU!’*

The official responsible for Gui and Fenghua’s area motioned for everyone keep moving. It seemed that an order to complete the evacuation as quickly as possible no matter what the city did had been issued and the soldiers were now actively pushing people toward the evacuation vehicles. Gui caught Fenghua’s eye as his segment of the crowd was broken off from hers and redirected toward a different vehicle. He signalled to her to follow, but she looked meaningfully up at the nearest screen, still displaying the record happiness index, and said, ‘I think I want to stay.’

He stared at her uncomprehendingly.

‘I don’t want to go back now’, she shouted across the widening gap between them in a way that suggested she was understanding something important about herself for the first time. ‘I worked so hard to get out of the countryside. You have your family in Chengdu. Your parents will be happy

for you to come home. The company will relocate—or you will find yourself a new job. You're already a manager!'

Gui's line was moving fast. He tried to break away and get back to Fenghua but a soldier grabbed him aggressively and shoved him in the other direction. Before he could understand what was going on, a crack erupted overhead leaving a weird visceral rumble in its wake, more felt than heard. Bursts of static leapt across the screens and the drones, convulsing like dying insects, started to tumble from the sky. The next thing he knew, he was speeding down the expressway, heading toward Lanzhou.

*

'Delusion clouds your sight.'

Gui squinted up at the figure standing over him. It was a young monk in a faded chogu and worn leather boots. The morning sun was now high in the sky and the monk was standing with it at his back. He shook a small silver bell in Gui's face.

'There is no shortcut to Enlightenment. You must journey through many lives. Lying flat traps the soul and prevents it from graduating.'

Gui made a pained expression and waved the monk away.

'You chase a deathless dream', he muttered to no one in particular as he continued on down the line, waggling the bell, 'even mountains crumble into dust.'

The temperature had risen considerably, and Gui clambered out of Snake's sleeping bag, gently massaging his stiff knee. When he stood up he could see that the queue, which followed the curve of the shell, now stretched all the way to the horizon. A drone branded with a state media logo passed overhead then circled back to alight on the roof of a nearby van, beside which a soberly dressed reporter was delivering a live broadcast.

'—was evacuated three years ago, when attempts to resolve the situation by decommissioning the City Brain core were unsuccessful. No longer home to the burgeoning tech culture that had given it life, Auspicious Nourishment was officially deemed a failed experiment and removed from the Pilot Project.'

The van's doors were open and Gui could see a broadcast technician and a drone operator sitting inside before a bank of monitors. The reporter

continued her story as the technician cut to a feed from a drone high above the city, revealing the black shell in its entirety.

‘However, recent inspections confirm that the city is still active and has been making alterations to its infrastructure, including erecting what appears to be a state-of-the-art solar energy dome, englobing the original city. LIDAR scans show an extensive underground complex housing a massive server farm and what experts believe to be a 0thorium nuclear reactor array. Auspicious Nourishment’s autonomous construction of these technologies is a testament to China’s global leadership in AI and a historic leap forward in clean energy innovation. For exhibiting such initiative, Auspicious Nourishment has been reinstated as the fifth star in the Five Stars Aligned City Brain Pilot Project.’

The drone that had landed on top of the van now rose and retraced the path it had been taking when Gui first spotted it flying low over the queue. The technician switched to the new feed.

‘These extraordinary achievements prove that the city was influenced only by the collective diligence and unwavering spirit of its original workforce, thoroughly dispelling the baseless rumours that the city had succumbed to the so-called doctrine of “lying flat”. The return of young, ambitious professionals, like those we see before us now, is a vivid symbol of the city’s bright future.’

Gui bowed his head as the drone passed over him. He didn’t want to be filmed. The triumphalist news report had succeeded in making him feel ashamed. Part of him wished it was true, and that Auspicious Nourishment had merely repaired and improved itself and that he was lining up to go back to his old job and his old apartment just a few doors down from Fenghua’s. But the real story was much weirder.

On the day Fenghua had stopped messaging him they had had an argument about her decision to return to Auspicious Nourishment. Like him, she had been evacuated the afternoon the kitten appeared. An assault that began almost immediately after the evacuation vehicles had departed and which lasted several days left Auspicious Nourishment as little more than a pile of twisted metal, rubble, and melted circuitry at the bottom of a crater. The debris was abandoned to the desert and the city removed from official maps.

But someone had made the mistake of assuming the city’s intelligence had remained centralised and all it had needed to regrow itself were a few sensors, some scalded batteries, and a couple of barely functioning drones.

Six months later, when Fenghua and a group of other ex-residents returned to explore the wreckage, they discovered that the city had been piecing itself back together. It spoke to them through its drones, calling itself ‘Chive City’, confirming Fenghua’s hypothesis, and Fenghua and a few others decided to return and lie flat with it.

Gui had called her a bourgeois individualist and begged her not to go. She’d retorted that the city wasn’t an individual and that if she became part of it, she wouldn’t be one either.

He’d told her she was selfish and a parasite on the social progress of others.

She’d called him a fuerdai and said he wouldn’t know.

That was the last time they’d talked. Fenghua wasn’t contactable after that. He’d sent her a message to apologise several weeks later, but she’d never read it.

Then the speculations started. It was said that the city had modified its old affective modelling software to create constructs of the lying flattists. That it had destroyed their bodies and uploaded them to its servers where they lived in a 3simulated recreation of the original city, or a permanent video game, or in some other virtual state of existence—as ghosts or spirits, immanent to the assemblage of electromagnetic waves, sensors, and probes. The stories varied.

By then Auspicious Nourishment had completed the construction of the photovoltaic shell, whose purpose was to act not only as a battery but also as a Faraday cage and shield. Nothing went in or out apart from the city’s drones and there was no way to know what was going on inside. A conspiracy theory circulated among employees of the companies that had once had their headquarters in Auspicious Nourishment that the lying flattists had effectively made themselves immortal. If the city had its own inexhaustible energy sources, if it was entirely autonomous, self-sufficient, and self-contained, what was stopping the lying flattists from living whatever lives they were living in there forever?

Ground down by three more years of not really knowing what he wanted or what he was supposed to do with his life, Gui had begun to think it might be a good idea to go back. It somehow seemed more reasonable to believe in the promise of immortality—and of fusing with Fenghua—than in the possibility of changing the conditions under which he lived his current life. So he had tendered his resignation, walked out on his wife, and caught the

train to Lanzhou. The taxi driver he had paid to take him to the middle of nowhere believed Gui was crazy until he saw the black shell appear on the horizon.

Not long after the media van and its drones had cleared off, Gui noticed a band of smaller, nimbler drones moving methodically around the edge of the shell, pausing every few metres to address the queue. He remembered this style of drone. They belonged to Auspicious Nourishment. In their wake, people were getting up, collecting whatever they had brought with them, and leaving. Others stayed put, glancing around defiantly. When the drones reached Gui he understood why. A synthetic voice was politely delivering the following message. '*WELCOME TO CHIVE CITY. PLEASE GO HOME. THERE WILL BE NO ENTRY INTO CHIVE CITY. SORRY TO BOTHER YOU!'*

The message was repeated, then the drones continued to fly in ragged formation down the line to inform the next group of people that they would not be let in.

Gui's phone chimed. It was a voice note from his wife. He pressed play.

'Gui, I saw a news report. What is going on? Are you in Gansu? Stop this nonsense and come home right now. Did you forget that you are a father?'

The phone chimed again.

'Who is going to look after your parents? And how am I supposed to pay the mortgage on my own?'

And again.

'Do you know how the city is powering itself? There are multiple nuclear reactors underground. It's dangerous! You need to leave.'

And again.

'I talked to your boss. He wants you to come back too.'

Li Gui rolled up the sleeping bag and crammed it with the rest of the jerky into the backpack Snake had sent. He opened a taxi app on his phone. For the first time in three years Auspicious Nourishment was on the map. He tried to book a ride but there weren't enough drivers nearby to respond to his request, so he opened a second app, hoping to maximise his chances. Auspicious Nourishment was on this map too. Sort of. A marker sat over his location but instead of 'Auspicious Nourishment' the characters read 'Chive City'. He switched back to the first app whose map now said 'Chive City' too. Just to make sure he wasn't going insane, Gui opened the Bountiful Harvest app. His location was marked 'Auspicious Nourishment',

but the characters were glitching in and out of legibility as if struggling against a force that was trying to transform them in real time. Then the page went blank and reloaded itself. ‘Chive City’ had won the battle.

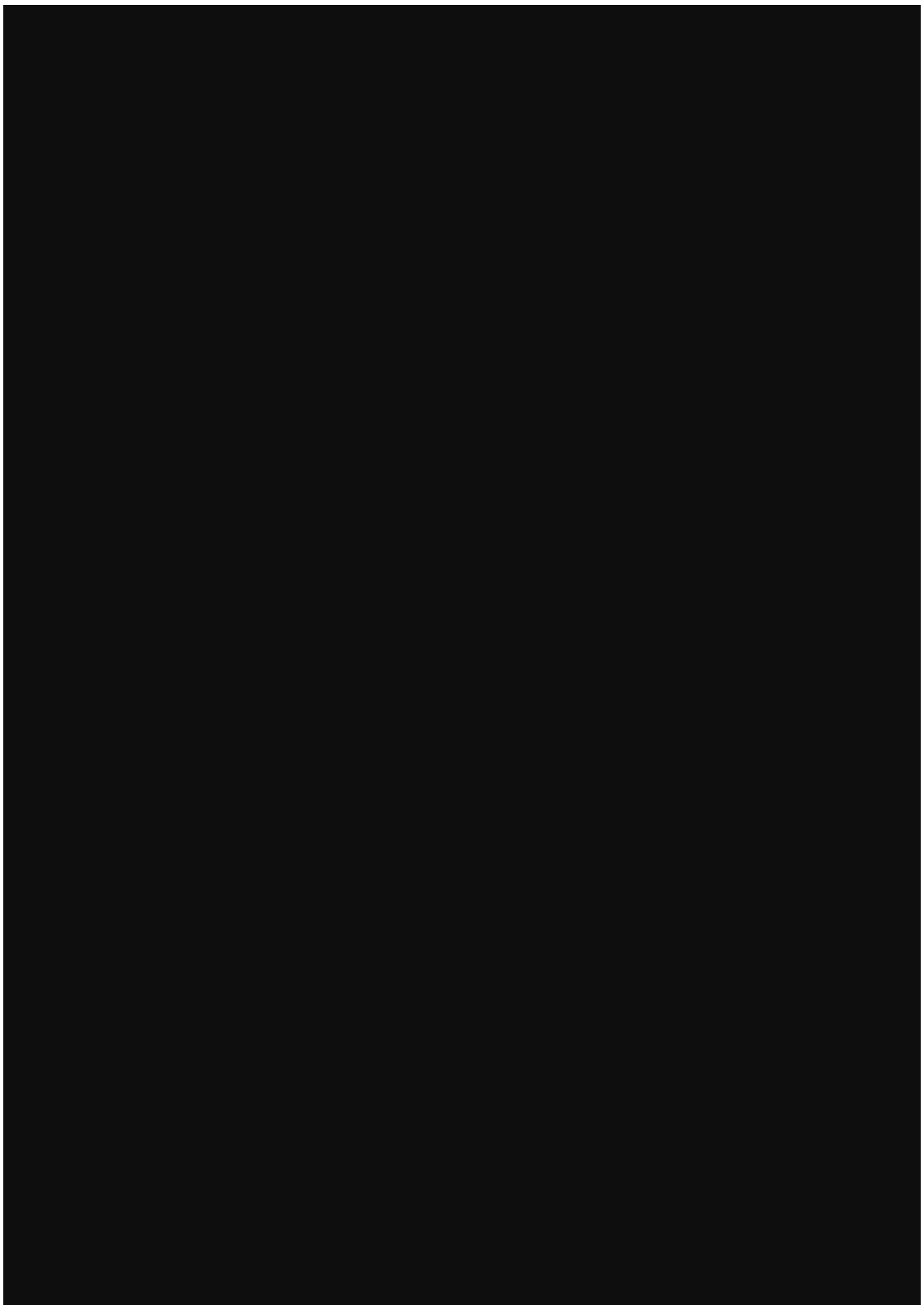
He turned around and gently placed his hand on the shell. It felt warm. ‘Goodbye’, he said, and he swore there was a pause before the shell started to flash its warning. His phone chimed again and a popup appeared, announcing that his taxi request had been accepted.

Tomorrow night he would sleep with his wife in their bed. The following morning he would wake up at six a.m., drink a glass of soy milk, and eat the cold steam bun she had left out for his breakfast. Then he’d make his way to the street, scan a QR code on a bike at the gates of the apartment complex and cycle to work, turning at the end of the road to join the other commuters, moving together in a swarm that never thinned out regardless of how many intersections it crossed, for each departing cyclist would immediately be replaced by another, emerging from a side road to join the infinitely pedalling, inexhaustible machine.

Notes

1. To be ‘harvested like chives’ is internet slang deriving from a long-running meme in which the image of garlic chives or *jiucai* (韭菜) being harvested with a sickle operates as a metaphor for ordinary citizens whose labour, energy, time, and/or finances are exploited by the state (or other powerful actor), which takes them to be an infinitely renewable and expendable resource, like chives. The particular version of the meme quoted above (‘Chives that lie flat are difficult to harvest’) has its origin in the now notorious post ‘Lying Flat is Justice’, *tangping ji shi zhengyi* (躺平即是正义) published in April 1 on the social media platform Baidu Tieba, which inaugurated the rise of *tangping* (躺平) or ‘lying flat’ as a social movement. The post was removed by censors shortly after it went viral. For an in-depth history and analysis of the chive meme, including in the context of *neijuanhua* (内卷) or ‘involution’, see Pang Laikwan, ‘China’s Post-Socialist Governmentality and the Garlic Chives Meme: Economic Sovereignty and Biopolitical Subjects’, *Theory, Culture & Society* 39:1 (2022), 81–100. On the memetic history of *tangping* and its relation to China’s 2021 Three-Child Policy see Zhu, Yin and Peng, Junqi, ‘From

Diaosi to Sang to Tangping: The Chinese DST Youth Subculture Online', *Global Storytelling: Journal of Digital and Moving Images* (2024) 3:2:13–38.



Benjamin Bratton

Afterword: Apology for Western Theory of AI

In early 2025, observers in China may have wondered why many Western AI-watchers were so surprised to see a Chinese AI company enter the rarified list of leading open weights foundation models. But this surprise was understandable: the ascendancy of DeepSeek called into question so many presumptive Western narratives about what AI in China is and may become. This was not Skynet AI, but Street AI: fast, cheap, good, and easy to use, and it made the supposedly ‘open’ Western way of doing things look downright bloated and locked down in comparison. Perhaps more importantly, the arrival of DeepSeek suggested that the steering of large models toward *open* multilingual futures might—via the hypercoupling of models, training techniques, inference subcultures, prompt cosmologies, and application wrappers and scaffolds—delay or reverse the apparently imminent decoupling of the West and China driven by their antagonistic geopolitical positioning. Just as life finds a way to evolve, so does technology—because technology is a kind of life, and it will follow an evolutionary path that is irreducible to its instrumentality for human ends. If artificial superintelligence kicks in sooner rather than later, replicating itself and terraforming its host planet toward full Kardashev I, then who patrols the South China Sea will probably be the least of our concerns. With the memorable DeepSeek panic in mind, my ‘apology’ to the Chinese reader is an attempt to explain *why* the response in the West was one of surprise and discombobulation. Below I will summarise what I see as some persistent Western habits of thought—habits that may seem odd because, seen from some distance, they are.

An inspiration for this book was the simple observation that no two cultures define ‘artificial’ and ‘intelligence’ in the same way and that therefore any two cultures’ respective understandings of ‘artificial intelligence’ are unlikely to be mutually transparent and aligned. During and after my stint as a visiting professor at New York University Shanghai,

this truism became less academic and far more geopolitical in tone. While conversation about what AI is should be planetary in scope and purpose, it is clear that not only are Western concepts of AI unnecessarily predominant, so are Western critiques of AI. Both focus on correspondences between AI and particular Western philosophical and political commitments in ways that may be locally productive but also globally limiting. As a Western theorist, I am disheartened by the genres of speculative critique that orient AI discourse not only because these approaches are provincial but because they are inherently flawed. Their overinvestment in individualisation, artifice, representation, narrow fictionalisation, and ultimately eschatology has established a standard repertoire with which to critically debunk AI, often projecting values onto it as a means of explaining it away, and doing so at the expense of approaches that would seek to generate theory from more direct encounters with the technology.

This afterword, and this volume as a whole, are not only about ‘AI and China’ but also about ‘AI and the West’, and in particular, about how the West imagines AI in China. Approaching the topic in this way means reframing the more troublesomely hegemonic forms of Western theory of AI and turning the critical lens onto the critique itself. If this has the effect of representing these forms of Western theory to Chinese readers in a way that underscores its limitations and undermines its false universality, then they will have achieved their goal. China and the ‘West’, especially the United States, conceived and composed AI together, but over time it became a wedge between them, and has come to define their intensifying antagonism. That antagonism is partly the result of misunderstandings related to how each imagines the other, and it is clear that better translations are required between different ways of defining what AI is and what it is for. Western theory of AI may presume that its axioms are more universal than they are. If so, they must be provincialised from the inside. This requires an apology (in both senses of the term).

Machine Decision is Not Final is a different kind of book about ‘China and AI’ from those in the growing pile of Western titles offering lurid—albeit not always inaccurate—tales of technological oppression, often pushing simplistic plotlines in which China is the ascendent cyborg geopower and the West the noble defender of the prerogatives of liberty. In the worst of these narratives, themes of ‘Oriental despotism’ return from the racist subconscious of well-educated columnists as a foil for freedom both

real and proposed. Concomitant with the original Romantic response to the rise of industrial capitalism in Europe, the philosophy of AI taught in the West today is one largely grounded in a moral suspicion of technology as a dehumanising force. Directly and indirectly, a fear of technology and its consequences for the planet and for what it means to be human sometimes mutate into a fear of China as a global agent of change and looming degradation. The most hardcore exponents of this view transform the 'yellow horde' of yore into waves of emotionless robots coming to rob them of their jobs and their dignity.

When examined closely, so much of Western theory of AI is not really what it thinks it is, and the figure of 'AI' with which it is preoccupied is not necessarily the AI right in front of us. AI is not all seamlessness and power; it is quite messy, unpredictable, slippery. Increasingly, the Western 'AI critic' is becoming little more than a cultural position to be adopted in opposition, not just to machine intelligence, but to the malign spectre of control in general. For, beyond the question of 'the machine', this is what looms darkest. Perhaps the indignation is less with AI than with how the weird realities it generates are undermining prior, overinvested political commitments. This now orthodox mode of critique, which establishes what is sayable, hearable, and understandable, is one means by which Western theory of AI has overpowered other possibilities not only for critique but for AI itself. In this way, ironically enough, the vernacular style of AI resistance native to intra-influencer feuds in a few select cities has become a global hegemon over the AI cultural imaginary.

So how to proceed? The answer, less grandiose than some might crave, is by the careful work of translation and attention to one another. Among the chapters of *Machine Decision is Not Final*, I read some works as emblematic of the Western approach of which I am suspicious, while others clarify how it might be reoriented, some do both, and some speak a refreshing new language. These latter articulate views on AI that begin and end with different premises. Given the speed at which AI innovations appear, to compile a book is to work at a comparatively glacial pace. But this provides its own crucial perspective. Over the period of time between the conception and final realisation of the book, technologies referred to as 'AI' have become dramatically less hypothetical. Regular people have interacted with multi-trillion parameter models for the first time. Some have contemplated their mobile phones and realised that effectively they now

held more ‘neurons’ in their palms than in their skulls. Reality also sometimes bites hard: while writing her chapter for this book, one of the Chinese authors declared that she had ‘lost faith in AI’. Having lived through the response to the COVID-19 pandemic, she felt utterly uninspired to write about ‘the future of AI’, especially in China, and ended up exploring her sense of unease instead. Many if not all of the contributors, myself included, have ambivalent feelings about the technology. But while it may appear that this ambivalence is directed at the same object, in truth, this is not the case. And since we cannot presume that we are always talking about the same thing when we evoke ‘artificial intelligence’, we must try to understand one another’s ambivalence on its own terms, not only in relation to our own. For comparative ambivalence, or so it seems to me, is a precondition for the viable planetary future of artificial intelligence.

What is Chinese Theory of AI?

The further you zoom out, the more technological determinism becomes a feasible historical perspective, and AI is no exception. However, technological determination is itself indeterminate; its contingencies arise from diverse initial conditions and particular cultural teleologies which can steer (and be steered by) the same force toward very different outcomes. Such is the case with AI in China and the West: from the very start, the paths have forked. As Anna Greenspan and Bogna Konior point out in their introduction, and as Shuang L. Frost discusses at length in her chapter on ‘Translating Chinese AI’, the connotations of the Chinese term *rengong zhineng* (人工智能) are closer to ‘humanmade intelligence’ than to the ‘artifice’—in the modern Western sense of ‘inauthenticity’—implied by the English term ‘artificial intelligence’. The history of AI in China is not well known in the West, even among AI scholars, and especially among AI critics, who, when they bother to consider China as a central concern at all, are often content to cycle through the same six or seven themes: oppression, surveillance, control, efficiency, threat, data capture, IP theft, etc.. These are all worthy topics, but the history is far richer and stranger than many have cared to consider. In all of its aspects, that history was and is a crucial part of the planetary history of AI. A more direct engagement places it in contexts of Chinese moral philosophy, economic development, and political institutions that are far more complex than those rote themes can

accommodate. It extends, to provide just one example among many, from the fraught history of cybernetics before and after the Cultural Revolution, to the relationship between cybernetics and industrial policy in the 1980s, to automation in the twenty-first century. Other threads lead back even deeper into the past.

The established Western history of science and technology in China concludes that its development was hindered by the absence of foundational theory. In the fourteenth century, Chinese science and technology was more advanced than in the West, and yet industrialisation was established in Europe. Joseph Needham's famous explanation for this 'paradox' was a cultural-determinist one. He suggested that the cultural principle of a discoverable 'rational order' did not exert a gravitational pull on Daoist thought (hence China's lack of 'foundational theory') and that Confucian values of sociopolitical stability discouraged disruptions in the means and mode of production.[1] Other explanations, however, have not leant as heavily on culture: unlike Europe, China did not incorporate vast resources from a 'New World' and was subject to what Mark Elvin has called a high level equilibrium trap, with the Chinese preindustrial economy reaching a balance between supply of labour and demand for its applications.[2] Greater levels of mechanisation were therefore not required because there were enough people to realise projects, or so the story goes. Assuming that there is some validity to the equilibrium trap hypothesis, the cultural factors described by Needham may themselves have been the expression of deeper economic-technological circumstances, rather than their cause. This would shed light on modern understandings of AI in China according to which the 'intelligence' subject to 'artificialisation' is defined in terms of coordinated organisation and collaborative production (each person, machine, and process plays a role in a greater whole). Importantly, this understanding of 'intelligence' does not primarily seek to affirm or disprove that a machine mirrors the workings of an individual human mind, but rather suggests that it relates to forces that shape the *coordination* of minds and bodies into common compositions and acts of worldmaking.

Contemporary Chinese philosophy of AI is sometimes oriented around the reconciliation of machine intelligence with the moral and spiritual teachings of Confucianism, Daoism, and Buddhism. Among the pressing concerns for this approach is where to locate AI, both real and hypothetical, in the architecture of *tiandiren* (天地人)—Heavens-Earth-Humans.[3] This

description of cosmic forces and the integrated relationality of all things implies a prescriptive morality whose goal is to ensure the harmony of those relations, leading to further questions: Should AI be addressed as part of an ethical order of 'things'—*wu* (物)—or in terms of familial relations and obligations? Could AI teach us to be more moral creatures, and must we teach it to be a moral creature as well? As an expression of a higher intelligence, will communion with AI allow for a deeper appreciation of the comprehensive integration of the universe? Is the 'automation of intelligence' more Confucian, Daoist, or Buddhist? Is it a matter of aligning action with norms and order, the expression of the forces that constitute being, or of the logical and harmonious organisation of society understood as a single coherent organism? Such questions demonstrate that, just as in the West, here AI is not only an instrumental technology, it is an existential one. Not only will these traditions of thought continue to inform the emergence of AI in China, but the gradual integration of machine intelligence into infrastructure, production, coordination, and symbolic expression will transform these philosophies in their turn.

Invention and Implementation

At present, a geotechnological conflict between China and America tends to frame the development of AI in both hemispheres. This is not the first time either of these great powers has steered the invention and implementation of computational infrastructure in relation to a geopolitical adversary. In China, the introduction of infrastructural computation, via a programmatic implementation of cybernetics from the 1950s up to the 1970s, closely tracked the nation's variously dependent and independent relations with the Soviet Union, while in America, confrontational competition with the Soviet Union fuelled the development of military applications and supplied the funding upon which the nation's computer industry would grow up. What makes the current relationship different is that America and China have together led the development of the AI technologies they are now trying to split apart into separate, but isomorphic software-hardware stacks.[4]

Over previous decades AI has been developed by China, America, Chinese working in America, Americans working in China, and Americans and Chinese working together, yet it has come to be seen by both powers as the territory upon which the battle lines between them must be drawn. No

one individual exemplified this co-constitution of computational innovation and the violent peculiarities of its politicised disentanglement more than Qian Xuesen. Known in China as one of the most influential and revered engineers of the late twentieth century, having had a hand in the nation's economic industrialisation, its early computer and missile programmes, and its satellite and space programmes, Qian was perhaps the person most responsible for bringing cybernetics and integrated systems engineering to the PRC. From where did he bring it? From Los Angeles, of course. The 'father of cybernetics' in China, among his many other accomplishments, was a cofounder of Caltech's Jet Propulsion Laboratory, before being deemed a security threat and placed under house arrest in California during the Second Red Scare. Qian used his five years of detention to write the first Chinese-language textbook on cybernetics, and was deported to China just in time for the Sino-Soviet split and the Cultural Revolution. Largely as a consequence of the split, in China cybernetics fell into disrepute and was forbidden throughout the Cultural Revolution up until its revival in the 1970s and 80s as part of Deng Xiaoping's modernisation plans. Those plans prioritised the mechanisation of production in, for example, the special economic zone in China's Pearl River Delta region, and would ultimately shift the global balance of economic power. As the twenty-first century unfolded, collaboration between America and China, once seen as mutually beneficial, came to be thought of as dangerous, and today, debates between the two nations concerning AI are posed in zero-sum terms.

Where does this place AI in contemporary China? *Xin si da faming* (新四大发明)—the 'Four New Great Inventions'—is a slogan first deployed by Chinese state media in 2017 which evokes and updates the 'Four Great Inventions' (*si da faming*, 四大发明)—gunpowder, the compass, paper, and printing—of ancient China. The 'Four New Inventions' are high-speed rail, e-commerce, mobile payments, and bike-sharing. It is worth noting that none of these were invented in China, but arguably all were first realised there on a societal scale. (Perhaps they should be called the 'Four New Implementations'?) This might lead one to draw a conclusion very different from the paradox associated with Needham, whereby preindustrial China invented many technologies but the political-cultural milieu was not able to socialise them as well as that of the West. When it comes to AI, might the case be instead that the West 'invents' what China will then implement into the social fabric? If so, is an equilibrium trap at work, only this time in the

West? Are established Western systems too hard to alter because of their reliance on decentralised markets and their cultural focus on the intelligence of individuals rather than the intelligence of coordination? If so, might the same pattern observed by Needham hold, but in reverse, for how AI will evolve in China and the West from novel technological fact to new social form? Or perhaps the evolution of AI will only make these distinctions less clear-cut. Either way, the decision as to whether to prioritise implementation over invention or vice versa implies a very specific picture of where 'intelligence' is to be found and who or what is doing the 'artificialising'. Most obviously, 'the state' does not imply the same institutional mandate in the West as it does in China—indeed, as Vincent Garton observes in his contribution to this volume, a directly corresponding concept does not even exist in Chinese political thought. The assignment of keeping a society harmonious according to foundational principles is not the same as that of defending the terms by which citizens act of their own legal accord. In ways that would make little sense in America, the state in China not only has a role to play in the organisation of AI, but in a more profound respect, as Garton suggests, perhaps the state *is* the primary AI. It is not, however, open source.

These are only a few general principles for contrasting Chinese and Western theory of AI and they are inevitably coloured by my own perspective as a Western theorist. Even when qualified in this way, it is important to note that these differences are open-ended, not metaphysical or essentialist. They revolve around at least three concerns that seem central to Chinese theory of AI: a necessary but uncertain question as to where and how to locate AI among the constellation and obligations of *wu* (物)—'things', the location of 'intelligence' in the coordinated and automatised patterns of productive composition, and the relation of AI to mandates of governance. Importantly, however, none of these questions maps symmetrically onto corollary approaches in the West, where different presumptions take centre stage.

Motifs of Mystification

Western theory of AI features some of the most inspired scientific and philosophical work of the century, overturning cherished beliefs in favour of stranger and more beautiful insights. Yet it can also tend to coalesce

around a number of flawed, recurring motifs, some of which I shall sketch out here, and which I have chosen for the following reasons. Firstly, because they relate to particular discourses that are, I believe, misleading, regardless of their origin; secondly, because they imagine themselves to be universal when in fact they are provincial, and thirdly, because, as a global export, they constitute a dangerously inadequate basis for a viable planetary theory of AI. They are as follows:

(1) The vexed and self-contradictory *distinction between natural and artificial intelligence*, and related questions as to how one does or should resemble the other. This itself is an expression of how troubled the differentiation between the natural and artificial is, and how it haunts the West more generally.

(2) The presumed definition of 'intelligence' as corresponding to *the experiential consciousness of individuals*, whence the positing of the ability or inability of AI to accomplish full simulation of such consciousness as the core criterion verifying its existence.

(3) The repeated citing of gaps between AI performance and *the everyday 'experience of experience'*, accompanied by appeals to *embodiment* as the foundation of phenomenological authenticity, with AI's supposed disembodiment seen as its fatal limitation.[5]

(4) The transformation, as AI shifts out of the lab and into the world, of anthropocentrism and anthropomorphism into social reductionism and *sociomorphism*—the belief that AI can only ever represent the power structure of society and/or the project of ensuring that AI aligns with broadly defined human values.

(5) The positioning of AI in the midst of political narratives that oppose *freedom and control on an eternal spectrum*, thus forcing the conception of its applications to align accordingly.

(6) The tendency of the relationship between real and hypothetical AI that has driven both Western theory of AI and Western AI technology to give rise to *slippages between fact and fiction* that complicate the effectiveness of criticism.

(7) Lastly, the locating of AI within eschatological forms of speculative thought, where it is approached through the interchangeable terms of salvation and apocalypse.

In what follows, I shall discuss these motifs in detail, endeavouring to highlight the places where they intersect, mutually support, or follow on from one another.

Is AI Artificial?

Perhaps the most critical concern for Western philosophy, beginning with Plato but taking its modern acute form in the Industrial Revolution, has been the differentiation of the true from the illusory, the real from the shadow, and the authentic from the counterfeit, a motif often interpreted in relation to technology. Along the way, the idea of 'nature' has been

construed to serve different purposes. It has been something to be revealed; a formal order against which one is measured; an irresistible gravitational force that would frustrate attempts to violate its laws, and, more recently, a pristine and preceding state of unmolested harmony in contrast with the spoiling interference of human industry. This interference has been called many things including ‘artificial’, which is meant less in the sense of an ‘artefact’ (the trace of intentional handiwork) than of ‘artifice’ (illusive deceitfulness)—although it is, most tellingly, etymologically derived from both. The West’s long-standing philosophical concern with the moral status of fakery and deception is then already at work in this interpretation of the artificial in negative terms, often against the ‘natural’—a project which has animated many Western philosophies of technology, some of the most famous of which were formulated as reactions against the cultural and emotional disruptions of technical modernity. Almost in self-defence, clearly artificial constructions such as political or economic systems have been motivated to testify that they are instead based on ‘laws of nature’ that stand, as put sarcastically by Marx (who derisively compared those who police the distinction and hierarchy between natural and artificial institutions to ‘theologians’), ‘independent of the influence of time’.[6] It should come as no surprise then that the designed and artefactual qualities of AI are referred to, over and over again, as evidence that AI is fake, deceptive, a hoax, a knock-off, or fundamentally inferior to the natural standard. The deep-rootedness of this distinction in Western theory of AI can hardly be overstated. In many ways it reveals why the topic of AI has been in the West a perennial source of intricate and profound philosophical reasoning: it seems to simultaneously reaffirm and undermine many of the tradition’s most important axioms.

Doubles and Shadows

As for ‘intelligence’, it is true that a global scientific inquiry into its parameters, especially over the last century, has made the term at once more precise and more heterogeneous and open-ended. One reads exhilarating comparative portraits of human and other animal intelligences, and even vegetal and cellular intelligences. Daniel Dennett, for example, marks sequential evolutionary transitions between ‘Darwinian’ intelligence, deriving from blind evolution; ‘Skinnerian’ intelligence, from learned

feedback; ‘Popperian’ intelligence, capable of foresight and simulation of hypotheticals, and, eventually, ‘Gregorian’ intelligence, involving use of the tools of symbolic reasoning and communication.[7] The very posing of the question to which his schema is an answer is possible only because the status of intelligence has been artificialised in ways that have put into question its generalisability. One can imagine many ways to situate AI within and against this evolutionary arc, but what is essential here is that Western theory of AI and Western theory of intelligence have not only been mutually informative; they are co-constitutive. The motif of artificiality as counterfeit revolves around the locating of intelligence in the conscious experience of the human individual. Put differently, while intelligence in the broad sense is actively imagined in far more colourful ways, the ‘natural’ intelligence to which AI is repeatedly compared seems to be defined by narrow phenomenological truisms.

On this view, the individuation of intelligence is located not just in the brain, or in neuronal architectures, but more expansively in the qualia of experience, especially those associated with intuition, emotion, and creativity. The volume of thinkpieces published over the last few years to defend this imaginary fortress from AI beggars belief. For these writers, the artificiality and machinic quality of AI implies that it is, by contrast, rational, procedural, affectless, and orderly. The latter qualities are not excluded from the commonsense idea of intelligence, but they cannot possibly *contain* the former—emotion cannot be derived from rationality—and thus machine intelligence, cold and rational as it is, cannot possibly reflect the depth and complexity of ‘true’, ‘natural’ intelligence. This is not a recent twist. From the very beginning of practical AI, machine intelligence was put to the test in various contexts in which it was called upon to imitate a human and would be granted recognition only if it could pass convincingly as one. In one well-known test’s mise en scène, a single human sits on one side of a partition (representing humanity in general), while on the other side is ‘the AI’, often given a proper name. The human proceeds to interrogate their invisible interlocutor. Are they talking to another human or is this thing their artificial double, a shadow, an imposter? The test subject succeeds when it can convince its human interrogator that it is *just like them*. This dramaturgy makes sense only because of (1) the strong opposition of the natural and the artificial (human and machine) as ultimate categories for intelligence and (2) the locating of intelligence in a

singular subject (represented by the example human) whose own thinking serves as the test's key criterion. Pitting a 'natural subject' against an 'artificial subject' is thereby the logical means by which a candidate intelligence is diagnosed and measured. For Western theory of AI, the question is less 'What relationship and responsibilities might one have to this other intelligence?' than 'How does it compare to me and at what point does it "pass" as me or fail to do so?'.

This double-standard (it *must* be like me/it can *never* be like me) has been reinforced, challenged, and manipulated across Western philosophy and its critique of AI as both a real and a hypothetical technology. In short, all the ways in which machine intelligence can be shown to be different from, and unrecognisable as, individual human experience align with all the ways in which AI is said to fail to be real, fail to be realised, or fail to be authentic.

The Embodiment Riposte

Many real and hypothetical forms of AI have also been defined by less cerebral notions of human intelligence, based in what I will call the embodiment critique. This critique affirms that, unlike humans or animals, AI is not embodied in the world but rather thinks within a closed mathematico-logical system of self-referential codes. That this critique would continue be posed so emphatically during decades marked by the evolution of advanced robotics is curious, to say nothing of the fact that the important shift from top-down to bottom-up approaches (including that of deep learning) that has been definitive for establishing the contemporary paradigm was accelerated precisely by challenges to formal logic as the primary method for artificialising intelligence.

Ironically, this emphasis on embodiment rarely pays attention to actual physical and functional similarities between animal and machine intelligence. These include, for example, correspondences between biological and nonbiological forms of stochastic prediction at the levels of neurons, neural networks, and cortical columns. Instead, the 'embodiment' vaunted in such critiques often refers to neither biological or abiological bodies as such but (once again relying on phenomenology as a foundation) to familiar subjective experiences of being embodied. If for some Daoist approaches 'intelligence' can be seen as an expression of the environment in which its embodiment is situated, Western theory of AI inverts this order.

For it, the world is produced by the experience of self-conscious beings, rather than the world producing beings which have experience of it. Which is to say that the language with which embodied thought thinks intelligently in the world is concurrent with the formation of that world. So much for materialism.

An expanded version of the embodiment critique sometimes presents itself as a critique of the dualist idea that mind and body are separate, yet does so only by reinscribing that separation more thoroughly. For this version of the critique, supposedly disembodied intelligence is, once more, strongly associated with abstraction and rationality, whereas embodied intelligence is associated with immediacy and intuition. The critique insists that the ideology of AI prioritises the former over the latter; it then responds by counter-prioritising the latter over the former. This leaves intact, and indeed affirms, the dubious idea that AI is essentially rational and abstract, while also suggesting that rationality and abstraction are somehow not also embodied forms of intelligence that emerge from specific material-semiotic complexities, consigning them instead to a place beyond everyday reality. Dualism is tenacious.

These genres of the Western theory of AI pursue a ‘God of the gaps’ style of argumentation. Deployed in theological apologies, this rhetorical device locates the agency of God in an increasingly shrinking fraction of reality as a means of defending His existence against the tendency of scientific explanation to increasingly demystify what used to be seen as God’s direct handiwork. Similarly, the realm of the ‘authentically human’ is said to be unavailable to other forms of intelligence. This vein of thinking organises a protest, in the name of human sovereignty, around all that computation may never be, idealising and elevating its own self-encapsulation. In Continental Philosophy and International Art Theory, a more erudite version of this coalesces into communal incantations on behalf of ‘the immeasurable’, ‘the unquantifiable’, ‘the ineffable’, ‘the incomputable’, ‘the incommensurate’, ‘the unnameable’, ‘the unrepresentable’, and so on. At stake are at least two outcomes for Western humanism. Firstly, as the list of things that machine intelligence is unable to accomplish grows smaller, the value of what remains grows in significance, and secondly, as that value grows it comes to further *constitute* the defensible definition of the ‘human’ as such. Not only are these values legitimately under siege by AI, but by association so too is the vulnerable core of humanity’s self-image. For Western thought, AI

comes to define the human not by reflection, but by way of negative theology.

Society as Essential Reflection: Must Be/Cannot Be

This existential crisis is aligned with the Western philosophical tradition in which ‘the question concerning technology’ finds technology to be, in its ‘essence’, fundamentally alienating for ‘Being’. This conservative recuperation of authenticity necessarily involves an aesthetic resistance to technology along with attempts to establish political autonomy from the latter’s ongoing incursion. Paradoxically, while so much of the Western philosophical discourse surrounding AI is invested in the conviction that AI *cannot ever be like us*, human experience as the ultimate measure of AI manifests itself in a Western political project that is just as strongly, but contrarily, aligned around the idea that AI *can only ever be like us*: a reflection of the inequitable society that made it. ‘All technology is human’, as one evangelist school of AI criticism puts it.[8]

The mental rotation taking place here moves from AI anthropomorphism to AI sociomorphism: from the premise that AI can be interpreted as a reflection of a human individual to the premise that AI can be interpreted as the reflection of human society—and is therefore the imprint of the hierarchies, divisions, biases, norms, and violences that characterise it. This popular (and often populist) approach takes forms that are frequently contradictory but resolvable if politically aligned around a militant autonomy from the perceived dehumanising incursion. For some, AI is a hoax (‘AI does not exist’) not only because it cannot replicate the primate mind, but because it is ‘just statistics’, weaponised by elites who deliberately or unwittingly deploy the technology in ways that freeze, replicate, reinscribe, and reinforce matrices of power. AI is identified by other critics as ‘Power’ itself (capitalised as in the Foucauldian usage), synthesised and incarnated against human autonomy and agency. For those for whom infrastructures are, in the last instance, expressions of cultural values, AI is, despite its deeply global origins, the singular cultural output of a particular region of Northern California. For some AI means the violence of the state, for others the avarice of corporations, for others the tyranny of the quantitative over the qualitative, for others the strategy of a particular hegemonic class, for others a gendered technology of desire and

division, and for others still, AI is the Northern Hemisphere and its entire colonial history compressed into one big signifier.

In this maximalist version of the sociomorphic critique, not only is the past placed under the dominion of the West, but so is the future, such that there is no escape from the Western imperium, now unhelpfully credited with extensive (imaginary) new powers. For example, the definition of 'extraction' is expanded to include not just the theft of natural resources by powerful actors, but any potential form of purposeful, 'foreign' artificialisation of the world that a critic might seek to delegitimise. The *production* of data—which by definition is not scarce—is directly conflated with the *extraction* of oil—which is. Governance through sensing and modelling is linked with the legacies of Western colonial violence which it not only resembles but purportedly exemplifies. 'Don't you see? It is *all* us. Everything!' Anything that suggests AI might not be sociomorphic to structures of Western power threatens to dilute the political commitment made through that conflation. It must lock in the negationist vision and disallow any entertaining of a future of AI that does not first settle accounts with History: a bar set so high as to effectively tether AI to infinite debt. This is epistemological capture presented as admission of guilt.

'Control Needs Time'

Because for proponents of the sociomorphic critique, AI already and inevitably reflects an illegitimate social structure, defending humanity (as a whole) against 'Power' dovetails with ensuring individual autonomy *from* AI. This cultural atmosphere has coloured the emergence of a profession called 'AI ethics', a field that includes a broad spectrum of serious and unserious research. Alongside nuanced analyses, initiatives under this umbrella also range from minimal compliance efforts for predicting possible harms in new products to wide-ranging intersectional critiques of capitalism and modernity, and sometimes all of these at once. In other contexts, AI ethics aligns with an earlier era of data politics and digital hygiene that focused on the protection of 'individual data' from large platforms and nosey states, the regulation of this data as a matter of civil rights, and the advocacy of tools for rendering oneself opaque. That is, the *individual* relationship to AI is still paramount to the extent that AI is framed in terms of command and control and the negative connotations

these cybernetic terms have for Western cultural norms across the political spectrum. When AI is identified not just with power structures but as *the power structure itself*, well-worn plotlines of sovereign individuals resisting being commanded or controlled can be counted on to spring up almost automatically, as if they had been lying in wait for the moment when they would once more be needed to give narrative shape to a confusing reality.

Reality Fictions

The blurring of fiction and reality, of hypothetical and hyperreal, has been an enormously important and productive path not only for open philosophical speculation but also for the practical development of fundamental technologies. AI, as a whole, is a technology that has always been developed in relation to thought experiments about AI. Its inventions have almost always first existed in fictional scenarios; in the West these have included scenarios such as interviewing an AI to discern its humanness; being locked in a room and forced to communicate in the pictograms of a foreign language; considering whether the machines at a mill possess mental states; auditing maximalist paperclip-making machines; contemplating how children deprived of sensory input might perceive colour; wondering on what grounds to prosecute the indecisive attendant of a deadly trolley car,[9] and more. Parallel to and sometimes intersecting with these scenarios is a rich legacy of literature and film that imagines the many guises of machine intelligence, both before and after its eventual emergence, as part of a vibrant and sometimes frightening world.

But the fluid passage between fact and fiction also contributes to an unfortunate tendency in the Western humanities to directly conflate textual representations of AI with AI itself, and the critical interpretation of the former with a political challenge to the latter. In this blurring of representation and represented, it is not uncommon for fictional or aspirational portrayals of AI (in TV and movies, or in corporate marketing media, for example) to be presented by able scholars in ways that imply that AI itself is being examined. Precariously, a methodological commitment to interpreting the world as a ‘text’ allows for the interpretation of texts as if they were the world. Work of this kind occupies an irresolvable gap between reality taken as fiction and fiction taken as reality, and risks turning into a fetish the failure of each to ever really fulfil the promise of the other.

Appreciating both the beauty and neuroses of these approaches sheds light on how AI is imagined in the West, as well as acting as a lens on the manner in which it is socialised, and demonstrating just how fundamental the question of 'what AI *is*' is to the West's core cultural battles. To the extent that AI is imagined as a civilisational technology, its implications can be absolute, transcendent, and/or apocalyptic. AI may usher in a new age of plenty or it may portend total extinction, or some unthinkable combination of the two. It may overtake humanity on some linear track of evolutionary intelligence or it may express and amplify the social structures of human societies, or both. It may constitute the foundation of a new transcendence or it may be the last stage of capitalism. From inside Western theory, it is sometimes difficult to separate realistic risk (of which there is plenty) from secularised eschatological panic.

This disposition manifests itself in calls for pushing imaginary stop buttons, bombing data centres before it is too late, putting public interest lawyers in charge of AI as soon as thermodynamically possible, righteously interpreting new technologies as incarnating the Mark of the Beast, laying bets on algorithmically driven superextinction events, not saying 'Roko's Basilisk' three times in the mirror, building secure containment rooms for chatbots, and so much more. The great American tradition of eschatological social panic does not need much time to appear when summoned. Future historians, whether human or AI, will also marvel at how the subcultures most ready to accept AI salvation so effortlessly rotated into those most willing to entertain AI damnation.

The Reflection's Reflection

Western theory of AI holds that AI, which is not at all natural, both does and does not exist and is both smarter than us and fundamentally and permanently more narrow-minded; it is undecided as to whether AI mirrors individual intelligence—except when it determines that it never really can do so (and therefore does not exist); it sees AI as disembodied virtual rationality and therefore as quite unlike humans, yet maintains that it is the direct expression of Western history and of its global reach; AI is both a force for the emancipation of individual intelligence into new realms and a

means for the social domination of naturally free individuals and communities, and therefore AI will either save the world or end the world.

Where does all this situate China, or other non-Western actors, in the project of constructing a planetary theory of AI? At the moment, the Western perspective allows only for specific contributions in which China is forced to play a limited set of roles if it wishes to enter into dialogue with the West: techno-despot or orientalised Other? Either way, China supposedly must act in the image of the West: in the first case because, according to Western theory of AI, AI is ontologically Western; in the second case because China's otherness is determined in advance by symmetrical opposition and complementary inversion—AI in China must be *all the things the West thinks of as the opposite of itself*. China is heard only when it speaks from a position of alterity (or subalterity), thus reaffirming the relation between the supposed hegemon and its subordinates. China is asked to speak, not on behalf of intelligence-as-such, but on behalf of particularity, resistance, experience, indigeneity, communitarianism, essentialism, timelessness, wisdom, and most of all, cosmology. Nothing could be further from the grounding upon which a viable planetary philosophy of AI might be collectively conceived, composed, and inhabited.

I hope I have provided, especially for Chinese readers, a high-resolution snapshot of a problematic fraction of Western theory of AI, and helped to lay bare some of its contradictions and inconsistencies. The ongoing formulation of a planetary theory and practice of artificialising intelligence in its diverse forms need not be beholden to such confusion. AI was formulated through a *global* history of abstraction, calculation, and speculation. That it grew through applications both profound and dreadful means it may not be so unlike natural intelligence after all. Comparative ambivalence toward AI is essential in understanding how initial conditions now being set may orient its evolution. This cannot be achieved by situating machine intelligence simply as an expression of human cultures, let alone of only one among them, but only by situating both human and machinic cultural forms within the twisting evolution of intelligence as a fully planetary process that is now staring back at us.

Notes

1. See J. Needham, *Science and Civilisation in China* (Cambridge: Cambridge University Press, 7 vols, 1956), vol. 2.
2. M. Elvin, T*he Pattern of the Chinese Past: A Social and Economic Interpretation* (Palo Alto, CA: Stanford University Press, 1973).
3. See Bing Song (ed.), *Intelligence and Wisdom: Artificial Intelligence Meets Chinese Philosophers* (Singapore: Springer Press, 2021).
4. See my 'On Hemispherical Stacks', in L. Dellanoce, A. Khalaf, K. Kuitenbrouwer, N. Nyabola, R. Roukens, A. Steiner, and M. You (eds.), *Vertical Atlas* (Rotterdam: Digital Earth, 2022).
5. Hubert Dreyfus offered a comparatively sophisticated version of this critique a half century ago. H. Dreyfus, *What Computers Can't Do* (New York: MIT Press, 1972).
6. K. Marx, 'The Poverty of Philosophy', in *Marx-Engels Collected Works* (New York: International Publishers, 50 vols., 1976), vol. 6, 105.
7. D.C. Dennett, *From Bacteria to Bach and Back: The Evolution of Minds* (New York: WW Norton & Company, 2017).
8. See *All Tech is Human*, <https://alltechishuman.org>.
9. The Turing Test; Searle's Chinese Room; Leibinz's Mill; Bostrom's Paperclip Maximiser; Jackson's 'Mary's Room', and The Red Trolley Problem respectively.

Notes on Contributors

Blaise Agüera y Arcas is a VP and Fellow at Google Research, where he leads an organisation working on basic research and applications in AI. He is the inventor of Federated Learning, a privacy-preserving approach to machine learning, and many AI features on Android phones were developed by his team. Blaise also founded Google's Artists and Machine Intelligence program and has published widely, including essays in *Noema* and *The Economist*, as well as authoring two books, *Ubi Sunt* (2022) and *Who Are We Now?* (2023). He is a recipient of MIT's TR35 prize and has given several TED talks.

Bo An holds a combined PhD in East Asian Languages and Literatures and Film and Media Studies from Yale University. Based on research into the history and theory of information technology in modern China, his dissertation examines the rise of computing in the People's Republic of China between the 1940s and 1980s. An is currently investigating post-1980s applied computing and the longer global history of science behind it.

Benjamin Bratton is a philosopher, design theorist, sociologist of technology and the author of *The Stack: On Software and Sovereignty* (2016), and most recently *The Revenge of the Real: Politics for a Post-Pandemic World* (2021). He is Programme Director of the New Normal at Strelka Institute in Moscow, Director of the AI & Culture Research Centre at NYU Shanghai, and Professor of Digital Design at the European Graduate School and of Visual Arts at the University of San Diego. <http://www.bratton.info/>.

Chen Qiufan (AKA Stanley Chan) is an award-winning Chinese speculative fiction author, translator, creative producer, and curator. He is Vice President of the Chinese Writers Association Science Fiction Committee, Research Scholar at the MacMillan Center of Yale University (2023–2024), Berggruen Institute Fellow (2023), Culture Leader of World Economic Forum (2018/2019), and has a seat on the Xprize Foundation Science Fiction Advisory Council. His works include the debut novel *Waste Tide*, as well as *AI2041: Ten Visions for Our Future*, co-authored with Dr. Kai-Fu Lee.

GABRIELE DE SETA is, technically, a sociologist. He holds a PhD from the Hong Kong Polytechnic University and was a Postdoctoral Fellow at the Institute of Ethnology, Academia Sinica in Taipei. Gabriele is currently a Postdoctoral Researcher at the University of Bergen, where he is part of the ERC-funded project 'Machine Vision in Everyday Life'. His research work, grounded on ethnographic engagement across multiple sites, focuses on digital media practices, sociotechnical entanglements and vernacular creativity in the Chinese-speaking world. He is also interested in experimental music, internet art, and collaborative intersections between anthropology and art practice.

SHUANG L. FROST is an Assistant Professor of Digital Innovation and Business Transformation in the Department of Digital Design and Information Studies at Aarhus University. She received her PhD in social anthropology from Harvard University in 2020. As an ethnographer, she has explored the evolution of ride-hailing platforms in China, the empowerment of Black women tech founders in Detroit, the history of taxi entrepreneurs in Republican era Shanghai, and more. She is currently leading a multi-year research project entitled, 'Cultivating Women Tech Founders in Denmark', with financial support from the Danish Independent Research Fund. Shuang lives on a farm on the Danish island of Samsø with her family and their chickens.

VINCENT GARTON is a software engineer, analyst, and independent researcher. His current interests focus on the intersection of aesthetics, technology, and political economy, as well as the generative possibilities of Chinese and continental philosophy. He previously studied political thought at Cambridge University.

STEVE GOODMAN (aka Kode9) is an electronic musician, artist, and writer based in South London. He founded the record labels Hyperdub in 2004 and Flatlines in 2019 and has released 5 albums and numerous mix compilations. His book *Sonic Warfare: Sound, Affect and the Ecology of Fear* (MIT Press) was published in 2009 and as a member of AUDINT, he co-edited *Unsound: Undead* (Urbanomic, 2019). His sound installations have appeared at Tate Modern, as part of the *AI: More than Human* exhibition at the Barbican in London, and CAC in Shanghai.

YVETTE GRANATA is a media artist and media scholar. She is Assistant Professor at University of Michigan in the department of Film, Television and Media and the Digital Studies Institute.

ANNA GREENSPAN is an Associate Professor of Contemporary Global Media and the Director of the Centre for AI and Culture at NYU Shanghai. She was a founding member of the Cybernetic Culture Research Unit at the University of Warwick, where she received her PhD in philosophy. Her most recent publication is *China and the Wireless Undertow: Media as Wave Philosophy* (Edinburgh Press 2023).

AMY IRELAND is a writer and theorist whose work focuses on gender and technology, and questions of human and machine agency in modernity. She is a member of the technomaterialist, transfeminist collective Laboria Cuboniks, whose *Xenofeminism: A Politics for Alienation* (2018) has been translated into 18 languages. An anthology of her writing in Spanish translation, *Filosofía-Ficción: Inteligencia Artificial, technología oculta y el fin de la humanidad*, was published by Holobionte Ediciones in 2022. With Maya B. Kronic she is the author of *Cute Accelerationism* (Urbanomic, 2024).

BOGNA KONIOR is an Assistant Professor of Media Theory at NYU Shanghai. Recently, she was a Research Fellow in the Antikythera Program on Speculative Computation at the Berggruen Institute, and a mentor in the Synthetic Intelligence program at Medialab-Matadero Madrid. She is the author of *The Dark Forest Theory of the Internet* (Polity, 2025).

VINCENT LÊ is a philosopher, PhD candidate at Monash University, and former researcher in The Terraforming programme. He has taught philosophy at Monash University, Deakin University, and The Melbourne School of Continental Philosophy. His writing can be found in *Hypatia*, *Cosmos and History*, *Identities*, and *Šum* among other publications. His work focuses on the history of the post-Kantian transcendental tradition in relation to the philosophy of artificial intelligence. <https://monash.academia.edu/VincentLe>.

LAWRENCE LEK is a filmmaker, musician, and artist who unifies diverse practices—architecture, gaming, video, music and fiction—into a continuously expanding Sinofuturist universe. Over the last decade, Lek has

incorporated vernacular media of his generation, such as video games and computer-generated animation, into site-specific installations and digital environments which he describes as ‘three-dimensional collages of found objects and situations.’ Often featuring interlocking narratives and the recurring figure of the wanderer, his work explores the myth of technological progress in an age of artificial intelligence and social change.

Lukáš Likavčan is a philosopher. His research focuses on philosophy of science & technology and environmental philosophy. He is a Global Perspective on Society Postdoctoral Fellow at NYU Shanghai, where he also collaborates with the Center for AI and Culture. Likavčan is an author of *Introduction to Comparative Planetology* (2019) and a member of the More-than-Planet Working Group at Waag Futurelab. <http://likavcan.com>.

Suzanne Livingston is a scholar, futurist consultant, and curator of the world-famous *AI: More Than Human* exhibition at the Barbican Centre in London. Her research concerns ideas of the human change under technological pressure and scientific paradigm shifts. She collaborated with companies and museums worldwide, including Sony, Reuters, Guggenheim, Whitney Museum and numerous others. <https://www.drsuzannelivingston.com/>.

Iris Long is a writer and independent curator whose research focuses on the megastructures of science and technology in China, as well as the psychogeography of technoscience. Her curatorial approach integrates science fiction, field trips, and deep engagement with scientific institutions. As a researcher, she has received fellowships and support from the Berggruen Institute, Swissnex, and others. Since 2023, she has been running the artist-in-residence program on art and astrophysics/astronomy at the Department of Astronomy, Tsinghua University. In 2021 she launched *Port: Under the Cloud*, a long-term research and curatorial project exploring the infrastructures of science and technology in China (<http://temporalstack.com>)—this is her passion project. Iris is also an amateur radio operator and a certified electrician. <http://irislong.xyz>.

Bingchun Meng is a Professor in the Department for Media and Communications at LSE. She is also the Co-Director of LSE-Fudan Global Public Policy Research Centre and Director of LSE PhD Academy. Her

research interests include gender and the media, political economy of media industries, communication governance, and comparative media studies. Her book *The Politics of Chinese Media: Consensus and Contestation* was published by Palgrave in early 2018. She is currently working on another monograph under contract with Columbia University Press about AI industries in China.

REZA NEGARESTANI is a philosopher and a head of the philosophy programme at the New Centre for Research and Practice. He is the author of *Intelligence and Spirit* (2018), an examination of the history of philosophy as a program for general artificial intelligence, and co-author of *Chronosis* (2021), an experimental graphic novel, which blends the genres of philosophy, science-fiction and theoretical physics. <https://www.urbanomic.com/contributor/reza-negarestani/>.

WANG HONGZHE was born in Anshan. He is now an associate professor in the School of Journalism and Communication, at Peking University. He has long been involved in advanced studies in media-centred humanities and social sciences. He graduated from the China University of Mining and Technology with a bachelor's degree in Materials Science and Engineering. He received his masters degree in Communication from Peking University, his PhD in Communication from the Chinese University of Hong Kong. His research interests include media history, Cold War history, information society, and labour studies.

REGINA KANYU WANG is a writer, researcher, and editor working in both Chinese and English. Her stories can be found in her individual collections *Of Cloud and Mist 2.2* and *The Seafood Restaurant*, various magazines, and anthologies. She is co-editor of *The Way Spring Arrives and Other Stories: A Collection of Chinese Science Fiction and Fantasy in Translation from a Visionary Team of Female and Nonbinary Creators*, *New Voices in Chinese Science Fiction*, and the English version of *The Making of The Wandering Earth: A Film Production Handbook*. She has won multiple Chinese Nebula Awards, and is a finalist of Hugo Awards and Locus Awards, for her stories, reviews, editorial, and other work.

XIN WANG is a curator and art historian based in New York. A PhD candidate in Art History at New York University, writing a dissertation on Soviet Hauntology, she held curatorial and educational positions at the

Metropolitan Museum of Art and the Whitney Museum of American Art, and received the Warhol Foundation's Arts Writers Grant in 1. Publications such as 'Asian Futurism and the Non-Other' have been widely translated and taught in university curriculums. She has served on jury panels for The Shed, the Creative Capital Grant, and Anonymous Was a Woman, as well as as a regular visiting critic at Yale University's MFA program in Photography. She is the curator of the 4th art and technology themed biennial program—entitled 'To Your Eternity'—at Beijing's Today Art Museum in fall 2023.

XIA JIA (AKA Wang Yao) is Professor and Head of the Chinese Literature Department at Xi 'an Jiaotong University. So far she has published an academic collection in contemporary Chinese science fiction studies, as well as one novel and three collections of speculative fiction. Eight of her stories have won the Galaxy Award, China's most prestigious science fiction award. Her first English collection *A Summer Beyond Your Reach: Stories* was published in 2020. She is also engaged in other science fiction related works, including translation, screenwriting, editing and creative writing teaching.

YOU MI is a curator and a researcher based at the Academy of Media Arts Cologne whose work explores the connection between ancient and futuristic technologies, focusing specifically on Eurasia and the Silk Road region. She teaches at the Academy of Media Arts in Cologne, and was one of the curators of the Shanghai Biennale 2020–2021. <https://aalto-fi.academia.edu/MiYou由宓>.